PLANT PERFORMANCE AND BEAN QUALITY OF COFFEE (*COFFEA ARABICA* L.) UNDER RUST (*HEMILEIA VASTATRIX* BERK. ET BR.) CONTROL AND FRUIT THINNING TREATMENTS

A Dissertation

by

FABIÁN ECHEVERRÍA BEIRUTE

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Seth C. Murray
Co-Chair of Committee,	Patricia Klein
Committee Members,	Chris Kerth
	Benoit Bertrand
Head of Department,	David D. Baltensperger

May 2018

Major Subject: Plant Breeding

ABSTRACT

The coffee (*Coffea arabica* L.) plant is known for its special beverage. However, genotype (G), environment (E), management (M), coupled with biotic and abiotic stresses, the post-harvest processing, and roasting can change sensory, chemistry, and gene expression profiles. We investigated how yield (abiotic stress) and coffee leaf rust disease (CLR) (*Hemileia vastatrix* Berk. et Br.) management affected the cup quality and gene expression in two susceptible varieties.

At the field level, CLR chemical control treatment had the highest effect in reduction of incidence (-12%) and severity (-27%). The hybrid's vigor improved tolerance to CLR (4 to 5% less disease) and increase in yield (~20%).

Cup quality was qualitatively and quantitatively measured in optimal ripened fruits by two sensory analyses (SCAA and WCR sensory lexicon), revealing that 10 of 70 attributes were significantly affected by the treatments without substantial changes in perception. Parallel analysis of the volatile fraction using SPME-GC/MS, revealed that 18 of 154 chemical volatile compounds changed their abundance according to the treatments. Remarkably, acetaminophen was found for the first time in roasted coffee and in higher concentrations under stress. Further study of the coffee green bean's volatiles revealed that the compounds related to fatty acids were increased under biotic stress.

Using immature and mature fruits, we were able to explore the transcriptome of both cultivars under stress and found an active oxidation process occurring in the cell

walls. A total of 471 gene ontology (GO) functional terms organized in 19 categories were associated with differentially expressed genes (DEGs) according to the treatments, cultivars, and maturity stages. Sixteen candidate genes for later validation were reported.

We also explored the leaf transcriptome under stress. We found a differential response of the cultivars under biotic stress revealed by 88 DEGs mediating qualitative or hypersensitive response (HR) and quantitative or systemic acquired resistance (SAR). Both differential-defense responses are hypothesized as the cause of changes in cup quality and tolerance to CLR stress.

This study is a first step in understanding the complexity of the physiological, metabolic, and molecular changes in genotype by management interactions for coffee production, useful for future cultivar improvement.

DEDICATION

I dedicate this dissertation to my wife Tatiana, my sons Daniel and Andrés, and our families for their support and sacrifices made during this journey.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Seth Murray, for all his patience, guidance, and clever ideas on how to analyze and interpret this research and breeding technologies. From his example, I obtained the tools to understand how to improve not just crops, but also ourselves, by doing more for others, while improving our professional skills with knowledge, dedication, and hard work.

To Dr. Patricia Klein, who offered all her laboratory facilities and knowledge to obtain, analyze, and write the results of this research. It is because of her (and her lab personnel) that I improved my writing and explanation skills, but also was able to understand many other molecular tools relevant to help the goal of plant breeding in coffee.

Thanks to Dr. Chris Kerth for his inspiration and vision. Without his (and laboratory assistants Alejandra Ochoa and Kayley Wall) help, the preparation and analysis of the volatile compounds would not have been possible. His knowledge was instrumental for understanding our data and ascertaining the complexity of the bean chemistry.

Dr. Benoit Bertrand was the precursor of this project and with all his experience, connections, and perspective, was able to escort “the ship to the shore”. As a coffee breeder, he has changed many dogmas, and thanks to his interdisciplinary and international approach, is creating new opportunities for producing the next generation of coffee cultivars and scientists.

I also want to thank Dr. Ronda Miller for her help on the sensory analysis. Without her and Blyhe Beaves help, none of the quality attributes of coffee would be discovered. I appreciate her comprehension and help to interpret the data, but also on the importance of the creation of a common language to describe such complex human/related perceptions.

I want to thank my laboratory coworkers: Dr. Robert Klein, (almost Dr.) Nolan Bentley and Elmer Roldan, Dr. Mahnaz Kiani, Dr. Nikhil Patil, BSc. Julie McCollum and BSc. Natalie Patterson for their help in finding answers and solutions for and beyond my dissertation as well as for the friendly work environment they created. Without their enjoyable every-day interactions, the frustrations and barriers would not have been transformed into alternative approaches.

Thanks also to the professors, my friends, colleagues, faculty, and staff in the Departments of Soil and Crop Sciences, Horticulture, and Food Science, the Laboratories of the Institute for Plant Genomics & Biotechnology and Quantitative Genetics & Maize Breeding for making my time at Texas A&M University a memorable experience.

Thanks to World Coffee Research and the Coffee Research and Education Center directors and personnel, especially Dr. Timothy Schilling, Dr. Leonardo Lombardini, and Siaska Castro, for believing in me and supporting my sponsorship and expenses.

I want to thank the Professors at the Instituto Tecnológico de Costa Rica (ITCR), especially to Dr. Miguel Rojas and colleagues who shared their laboratories to allow me to conduct part of the research. I extend my gratitude to Professors MSc. Alberto

Camero, MSc. Sergio Torres, MSc. Wayner Montero, MSc. Ingrid Varela for their motivation and encouragement. I also extend thanks to the administrative personnel of ITCR, for their kind support for part of my family living expenses.

I want to also thank to Drs. Irena Hilje and Ramón Molina at the Universidad Nacional (UNA, Costa Rica), for allowing me to borrow their laboratory and equipment to dehydrate RNA samples.

Finally, I want to thank all the personnel at the Aquiares Farm, especially to Mr. Alfonso Robelo, MBA. Diego Robelo, Ing. Luis Guillermo Ramírez, Roger Solano, and Manuel Jara; MSc. William Solano at CATIE, and Dr. Fred Geoget from CIRAD, for all their collaboration at the farm. Their help was crucial for all the field research conducted for this study.

To all the ones I mentioned above and many others that didn't mention: Thank you!

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Dr. Seth Murray [chair] of the Department of Soil and Crop Sciences, Dr. Patricia Klein [co-chair] of the Department of Horticultural Sciences, Dr. Chris Kerth of the Department of Animal Science, and Dr. Benoit Bertrand from the World Coffee Research.

The sensory and chemistry data analyzed for Chapter II was provided by Dr. Rhonda Miller and Dr. Chris Kerth, respectively, of the Department of Animal Science. Dr. Miller had the corresponding authorization statement IRB 2015-0423M for the human based sensory evaluation. The data related to the volatile information in Chapter III was conducted by Dr. Chris Kerth and myself.

The roasting and SCA sensory evaluation was sub-contracted to Songer & Associates, Inc. (Boulder, CO). RNA-seq library preparation and sequencing was sub-contracted to Polar Genomics LLC (Ithaca, NY). All further work conducted for the dissertation was completed by the student independently.

This work was made possible in part by the United States Agency for International Development (USAID, APS-OAA-13-000003) and World Coffee Research funding.

Further scholarship for living expenses was also covered by the Instituto Tecnólogico de Costa Rica (ITCR).

NOMENCLATURE

CLR	Coffee leaf rust
CBB	Coffee berry borer
P	Phenotype
G	Genotype / Cultivar
E	Environment
M	Management
NGS	Next generation sequencing
QTL	Quantitative trait loci
eQTL	Expression quantitative trait loci
GWAS	Genome-wide association study
SNP	Single nucleotide polymorphism
GO	Gene ontology
GE	Gene expression
DEGs	Differentially expressed genes
JA	Jasmonic acid
SA	Salicylic acid
TF	Transcription factor
ROS	Reactive oxygen species
Redox	Reduction and oxidation process
NO	Nitric oxide

H_2O_2	Hydrogen peroxide
HR	Hypersensitive response
SAR	Systemic acquired resistance
RGA	Resistance gene analog
SH	Susceptible to *Hemileia*
Avr	Avirulence
LRR	Leucine rich repeats
C	No control treatment
R	Rust control treatment
T	Fruit thinning treatment
R + T	Rust control and fruit thinning treatment
TAMU	Texas A&M University
ITCR	Instituto Tecnológico de Costa Rica
UNA	Universidad Nacional
CIB	Centro de Investigaciones en Biotecnología
WCR	World Coffee Research
SCAA	Specialty Coffee Association of America
SPME	Solid phase micro extraction
GC-MS	Gas chromatography-mass spectrometry
NIRS	Near infrared reflectance spectroscopy
NMR	Nuclear magnetic resonance
HPLC	High-performance liquid chromatography

ANOVA	Analysis of variance
MANOVA	Multiple way analysis of variance
LSM	Least square means
LSD	Least significance difference
R^2	Coefficient of determination
SE	Standard error
RMSE	Root mean square error
FDR	False discovery rate
FC	Fold change
Q	Quantile normalization
PCA	Principal component analysis
SEA	Singular enrichment analysis
REML	Restricted maximum likelihood
LOX	Lipoxygenases
CYP	Cytochrome P450’s
13-LOX	Linoleate (13) S-lipoxygenase
PP2C	Probable protein phosphatase 2C 74
MAPK	Mitogen-activated protein kinase
ABA	Abscisic acid
DAP	Days after pollination
TAG	Triacylglycerols
CWP	Cell wall polysaccharides

CGA	Chlorogenic acids
XTH	Xyloglucan endotransglucosylase/hydrolase proteins
PAL	Phenylalanine ammonia-lyase
PDH	Pyruvate dehydrogenase
KASII	Ketoacyl-ACP synthase
FATb	Acyl-ACP thioesterase
ROI	Reactive oxygen intermediates
ETI	Effector triggered immunity
NAD	Nicotinamide adenine dinucleotide
TCA	Tricarboxylic acid
ATP	Adenosine triphosphate
OMT	*O*-methyltransferases
SAM	S-adenosyl methionine
ACC	1-aminocyclopropane-1-carboxylate
ACO	ACC oxidases
ETR	Ethylene receptors
SUS	Sucrose synthases
RFO	Raffinose biosynthesis-related
MAN-B	(1‑4)‑beta‑mannan endohydrolase
TE	Transposable elements
m.a.s.l	Meters above sea level
Ha	Hectares

TL	Total number of leaves
RI	Rust incidence
RS	Rust severity
RE	Rust sporulation
OC	Overall condition
TH	Total harvest / yield
S19	Cardboard aroma
S33	Burnt flavor
S26	Overall impact flavor
S29	Astringent flavor
S24	Stale aroma
V5	1,2,3,4-tetrahydronaphthalene-5-carboxaldehyde
V82	Acetaminophen

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Among developing countries, coffee is one of the most traded commodities. More than 120 million families are relying economically on the growth and exportation of this unique product. The United States of America alone imported an equivalent of $30.5 billion of coffee in 2015 [1]. Even when an increase in demand persists, during the last decade different economic and environmental conditions affected the reliability of the supply chain. Examples of the barriers include global economic crises, diseases, and climate change, which increased the cost of production and decreased the yield.

Climate change and actual agronomic practices have exacerbated abiotic and biotic stresses, which are expected to affect plant performance [2]. Some of the major problems in coffee production are the biotic stresses caused by insects, bacteria, and fungi, since these are related to environmental and management synergies. The most relevant disease on coffee is coffee leaf rust (CLR), caused by *Hemileia vastatrix* Berk. et Br., which is prevalent in tropical and subtropical regions where the best quality of coffee is produced [3-5]. To mitigate the problem and develop new resistant varieties with improved market quality profiles, more research is needed for understanding how coffee performance and quality is affected by CLR disease [6].

1.1.2 Origin and genetics of coffee

Botanically, coffee belongs to the *Coffea* genus and is one of 500 genera documented in the Rubiaceae family (the fourth largest family of angiosperms). The family belongs to the order Gentianales, which diverged around 87 million year ago (Myr) from Asterids and 120 Myr from Rosids [7]. When comparing coffee to the last common ancestor, *Solanum lycopersicum*, in terms of macrosynteny (co-localization of genetic loci on the same chromosome within an individual or species), coffee shares several features, including a similar genome size (704 Mbp), basic chromosome number (11), cytogenetic architecture, relative recent polyploidization, and expressed gene repertoires in the seed and the cherry. However, extensive microsyntenic rearrangements in coffee have occurred, leading to chromosomal evolution since their divergence [8].

As part of coffee evolution, more than 124 diploid species have arisen in the *Coffea* genus [7, 9]. Between all the species described within the genus, only two have economic relevance worldwide: *Coffea arabica* (Arabica) and *C. canephora* (Robusta). Arabica represents the gourmet coffee that ~65% of the world consumes and it is the only allotetraploid of the genera. *C. canephora* represents almost all the remaining global production (~35%) and is used in blends with *C. arabica* or for caffeine extraction.

The *Coffea* species were divided by Chevalier [10] into four groups: Eucoffea K. Schum. (true coffee), Mascaracoffea I Chev. (lower caffeine content), Argocoffea Pierre (today considered to belong to *Argocoffeopsis* Lebrun genus) and Paracoffea Miq (represented by the *Psilanthus* Hook genus). The first three groups include coffees

exclusively native to Africa, but most of the representatives of the fourth group are native to India, Indochina, Ceylon, and Malaya [11-12]. The Eucoffea section is the most economically important of the genus and can be subdivided into the following five subgroups: Erythrocoffea, comprising the species *C. arabica* L., *C. canephora* Pierre, and *C. congensis* Froehner; Pachycoffea, including *C. liberica* Hiern, *C. dewevrei* de Wild. et Dur., *C. klainii* Pierre, *C. abeokutae* Cramer, and *C. oyemensis* Chev.; Nanocoffea, with *C. humilis* Chev, *C. brevipes*, and *C. togoensis*; Melanocoffea, including *C. carrissoi* Chev., *C. stenophyla* G. Don, and *C. mayombensis*; and Mozambicoffea, which comprises species like *C. racemosa* Lour., *C. salvatrix* Swyn. et Phil., and *C. eugenioides* Moore [13].

The natural habitats and distribution of *Coffea* species in their center of diversity in Africa and Asia was under several arrays of ecological microniches. The evolution of *Coffea* varied from highlands to lowlands, drier to humid, warmer to cooler, and shared environments with a diversity of flora and fauna [14]. Even while the genus occupied wide and diversified environmental conditions, all diploid species maintained a similar chromosomal structure, which would arise from the same basic *A* genome (monophyletic origin) [15]. Apparently, effective geographical isolation has produced phenotypically divergent populations through genetic drift and natural selection pressure. For example, comparing Occidental and Oriental continental African species, the tree architecture, leaf shape and size, fruit size and color, and ripening time can be contrasted (Figure 1).

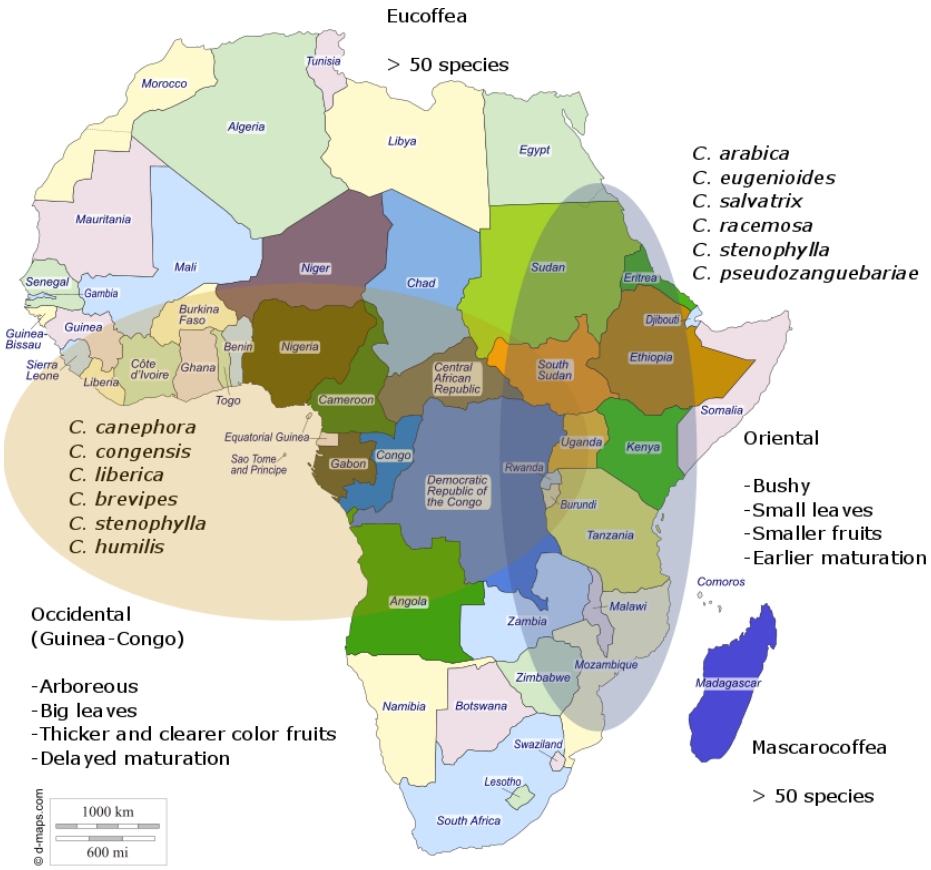


Figure 1. Biogeographical distribution of *Coffea* in the African continent.
African map modified from d-maps (http://d-maps.com/m/africa/afrique/afrique39.svg) in Inkscape 0.92.1 (Inkscape, Nederlands, https://inkscape.org/) software.

The regions where coffee is grown in America and Asia, represent a significant bottleneck from the origin in Africa. Considering the major species grown (*Coffea arabica*), the center of origin was established to be between the southwestern highlands of Abyssinia, Ethiopia, and the Boma plateau of Sudan. Historically, the Muslims during the 14th century, transported seeds to Arabia and also Yemen, which began coffee's domestication, cultivation, and commercialization [16]. It is believed that the common name "coffee" derived from the similarity to "qahwah," which means wine in Arabic.

During the 17th century, the crop was introduced to India. By the 18th century, the French were growing coffee commercially in Surinam and Martinique. In 1706, one plant (*C. arabica* var. Typica) was transported to the Amsterdam Botanical Garden (Holland), and in 1727, some seeds were taken from it and introduced to Brazil and later to other parts of America. Thus *C. arabica* var. Typica was the primary cultivar from which the first plantations were grown [17]. Between 1860 to 1870, seeds of *C. arabica* var. Bourbon, from the Bourbon Island (La Reunion), were also used to establish new plantations in Brazil. This type of coffee was the origin of the most relevant cultivars grown around the world, especially those derived from dwarf mutants including Caturra and Catuai. Even though the cup quality and adaptability of the Bourbon-derived cultivars is commercially accepted, these are susceptible to the most predominant leaf rust disease races [18]. The narrow genetic diversity of the *C. arabica* commercially grown is therefore increasing the risks of abiotic and biotic stress predisposition.

Genetically, *C. arabica* is an allotetraploid (2n=4x=44) that originated from an interspecific natural cross of ancestors closest to *C. canephora* and *C. eugenioides* (both diploids with 2n=2x=22) between 10-450 thousand years ago [19-20]. Besides *C. arabica*, *C. heterocalyx*, and *C. anthonyi*, the rest of the species have gametophytic auto-incompatibility. *C. arabica* behaves genetically as an interspecific hybrid due to its homoeologous genome constitution: E^a (*C. eugenioides*) and C^a (*C. canephora*) [21].

The decreased coffee population size and increased spatial isolation from wild to cultivated plants may be causing inbreeding at the expense of improving local fitness, associated disease resistance, productivity, and plasticity. The latest genetic diversity

study on cultivated and wild *C. arabica* plants (P. Klein et. al., [22], unpublished), found little diversity on the 846 accessions from the Tropical Agricultural Research and Higher Education Center (CATIE) (one of the largest and more representative germplasm collections *ex situ*). Since the study used a whole-genome deoxyribonucleic acid (DNA) re-sequencing approach, the results clearly indicated that there is a narrow genetic base from which to create new variability within the species. These results were also similar to studies conducted with RAPD, AFLP, and SSR markers on commercially cultivated plants [23]. The very narrow genetic base of the commercial coffee cultivars utilized today, the high susceptibility of commercial plantations to pests and diseases, global climate change with rising temperatures, and the overall monoculture agronomic practice used (like in almost all crops), represent major challenges for coffee production in the coming years [24-25].

1.1.3 Coffee leaf rust disease

Coffee plantations may be affected by several plagues and diseases that attack roots, stems, leaves, and fruits. The most common and widespread is "Roya" or coffee leaf rust (CLR), caused by the basidiomycete fungus *Hemileia vastatrix* Berk. et Br. (Figure 2) and to a lesser extent by *H. coffeicola* MaubI. & Rog. Coffee leaf rust (CLR) affects almost all *Coffea* species as well as the related *Psilanthus* genus. All over the world, CLR is considered the most important disease affecting coffee due its economic impact [26-27].

CLR was discovered in 1861 near Lake Victoria (East Africa) on wild *Coffea* species and described as *Hemileia vastatrix* by Berkeley and Broome in 1869 [28]. Around 1888, CLR destroyed coffee production in Ceylon (the most important producer at that time). In later years, CLR became widespread in Asia, Africa, and America; the last was first detected in Brazil in 1970. In Costa Rica, the disease was discovered in San Carlos in 1983, after battling more than seven years of agronomic practices to repeal the entrance from the other Central American countries [29].

The life cycles and genetics of CLR are complex and not completely understood. Similar to many other host obligate (biotroph) fungi, CLR is one complex rust from the Pucciniales order (Basidiomycota), the largest group of fungal plant pathogens [30]. Molecular studies revealed that the genus *Hemileia* is one of the most ancient rust lineages of the order and actually included in the Mikronegeriaceae (Cummins & Hiratsuka) family (characterized by a teliospore ephemera, thin wall and hyaline), which diverged 91-96 Myr [28]. The average genome size of the different species of the Pucciniales was determined to be around 225.3 million base pairs (Mbp), which is significantly larger than the average size of 49.9 Mbp within the Basidiomycota order. Even more striking was the fact that the CLR genome size ranks as one of the largest of the Pucciniales, with an estimated 733 Mbp [30]. Relationships between genome size and biological parameters have been linked to the ability of an organism to overcome selection pressure, thus indicating probable natural evolution and selection between and within hosts, increasing pathogenesis complexity [31-32].

CLR is a hemicyclic fungus producing three types of spores during its life cycle: urediniospores (dikaryotic cells that can only infect its host), teliospores (dikaryotic cells produced at the end of the infection cycle for survival purposes), and basidiospores (meiospores that supposedly infect an alternative unknown host) [30, 33]. The urediniosporic phase is the most important way of dispersion, survival, and inoculum of the fungus on coffee. Morphologically, the urediniospores are reniform, with a strongly warted convex face but smooth concave face. The wall is thick and hyaline, letting the carotenoid lipid guttules show the characteristic yellow-orange color [34]. The urediniospores are attached to the spherical, smooth, and hyaline sori in a bouquet shaped conformation [28]. CLR teliospores are commonly found between the urediniospores and seem to be derived from the same fertile cells in the sub stomatal chamber as urediniospores, but at the culmination of the urediniospore formation under dry and cold conditions [35-36]. From the cylindrical promycelium of the teliospore, one to four basidiospores are formed. Even though no alternative host has been found to date for CLR, it is still assumed that it is heteroecious (parasitic on different and often unrelated species of host at different stages of life). An alternative hypothesis is that the fungus is a primitive autoecious rust lacking pycnial and aerial stages of life cycle [33].

A B C

Figure 2. Symptoms of the coffee leaf rust (CLR).
A: Defoliation in a susceptible plant (right) versus a resistant plant (left); B and C: Leaf lesions on the abaxial leaf surface.

Similar to other fungi, CLR has different phases during the infection of the host (Figure 3), and some paralogue genes have been shown to interact according to the infection stage (Table 1). The first stage in the infection process involves specific events including appressorium formation over stomata (on the abaxial side of the leaf), penetration into the host leaf, and later colonization of living host cells by intracellular specialized fungal structures (haustoria). Next, the colonization of the mesophyll by intercellular hyphae will result in the production of asexual or cryptosexual (abnormal germination and asynchronous nuclear division) reproductive structures in the sori that will arise from the lesion. Every lesion can produce ~300,000 to 400,000 urediniospores. New spores can appear 15 to 24 days after infection under optimal field conditions [34, 37].

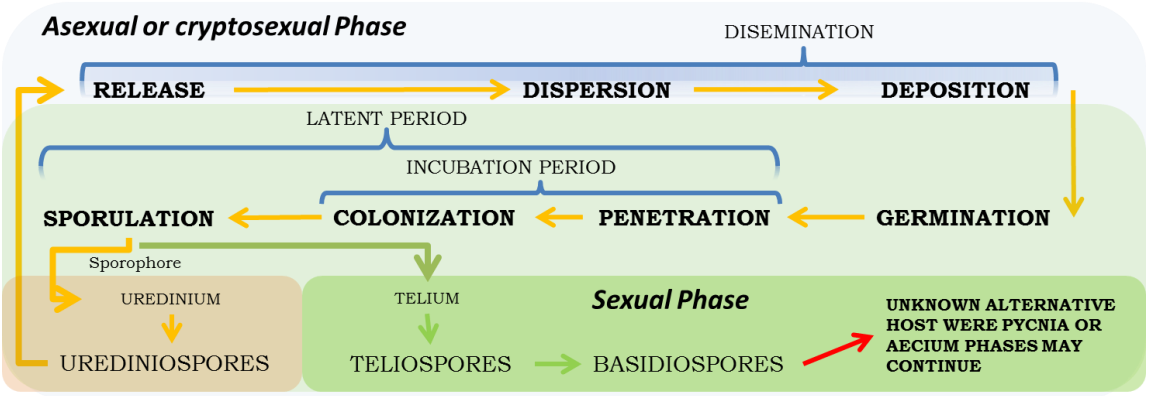


Figure 3. Schematic heterocyclic life cycle of *Hemileia vastatrix* Berk. et Br.

Table 1. Gene expression during rust infection in coffee.

Phase	Time	Structures	Expressed Genes
Germination	18 hpi*	Hyphae and appressoria formation	*Gp*-a, *MAPK*, *CytB*, *AAT3*, *HESP-379* and *Hv00571*
Penetration	48 hpi*	Hyphae stage, haustoria formed in stomata	
Colonization	7 dpi**	Haustoria in mesophyll, differentiation	*HESP-379*, *Hv00147*, *Hv00162*, *Hv00628*
Sporulation	21 dpi**	Extended sporulation	*CD1*, *MAD* and *RTP*

Created from: Fernandez, et al. [38].* hpi = hours post infection; **dpi = days post infection.

Because CLR has coevolved with coffee, it has been able to suppress host recognition or defense mechanisms, resulting in the development of 53 (or more) different physiological races according to the resistance genes developed by the plant. With such a high number of physiological races characterized and unknown sexual cycle in an alternative host, the variability of host resistance is expected to also be wide and complex [39].

1.1.4 Genetic resistance to CLR disease

Genetic resistance for several of the 53 races of the fungus has been found in coffee. Nine monogenic dominant resistance (*R*) genes, denoted as "S_H" (Susceptible to *Hemileia*), have been discovered in *Coffea,* which confer vertical resistance (single gene in the host has a major effect in resistance to a pathogen race-specific gene) to the nine avirulence or *"avr"* genes from the pathogen [40-42]. From the nine proposed *R* genes in *Coffea,* S_H1, S_H2, S_H4, and S_H5 have been found in *C. arabica*, S_H6-S_H9 in *C. canephora,* and S_H3 in *C. liberica* [43]. A new $S_H?$ gene coming from *C. canephora* (or another species) is assumed to exist [44-46]. Additionally, more than 40 candidate genes similar to *R* genes and defense response homologues to ones in the genomes of other organisms have been found which are related to partial resistance.

The resistance or susceptibility of coffee plants to each CLR race is determined by artificial inoculation [40]. Validation is performed by infecting leaf tissue from each tested genotype with already isolated and characterized CLR races. According to the disease reaction (i.e., resistance, tolerance or susceptibility), the plant is classified into a physiological group of complete recognition profiles with the specific CLR race, named arbitrarily after the letters of the Roman and Greek alphabets. The races of CLR on the other hand, were classified *a priori* according to their effectiveness to infect a certain coffee plant with known genetic composition [39]. A subset list of the 17-21 plants with known physiological classification and races that are compatible are summarized in Table 2. Within each classification group, some resistant and some susceptible plants have been found, revealing the complexity of disease resistance to this pathogen.

Additionally the genetic complexity of the CLR races and number of races present at the same time infecting a single lesion are factors that influence effective host resistance, since this can generate increased selection pressure during several CLR life cycles and influence the generation of more complex races.

Table 2. Host physiological groups with compatible interaction with CLR races.

Host physiological group	**Compatible *H. vastatrix* races**	**Compatible *avr* genes**	**Coffee varieties within the group**
E, D, C, J, H, G	I, II, III, XV, VIII, VII	*S_H2, 3, 5; S_H4, 5*	Bourbon, Geisha, K7, SL6, Agaro, Kawisari
C, α, β, Ɣ, J, I, W	I, III, XV, XIX	*S_H1, 4, 5; S_H1, 2, 5*	Matari, Dilla & Alghe, S12
L, Z, W, H, Ɣ, X	I, III, VII, VIII, X, XII, XIV, XV, XVI, IV, VI, XI, XXI, XX	*S_H1, 4, 5; S_H1, 2, 5*	Kaffa, K. P
A, 1, 2, 3, 4, R	XXII, XXV, XXVI, XXVIII, XXIX, XXX, XXXI	*S_H3, 5*	Tymor Hybrids such as 832/2, T5296, S288
F, N, B, K, P, Q			*C. canephora, C.congensis, C.excelsa, C.racemosa*
O, S, V, X, Y, Z, 4-11			Synthetized by Centro de Investigação das Ferrugens do Cafeeiro (CIFC)

The compatible susceptibility of coffee to CLR is explained by a gene-for-gene interaction, where a specific *R* gene (*S_H1-S_H9*) in the plant recognizes a specific *avr* gene (*avr*1-9) in the pathogen that leads to a compatible recognition, which results in an infection. Resistance to CLR is obtained when no complete recognition is detected, normally by a combination of several *R* genes in the plant. Since 1911, the major breeding programs around the world have been releasing varieties with more durable

combinations of *R* genes with other few genes in order to prevent the compatible reaction of the CLR races with the host (Figure 4).

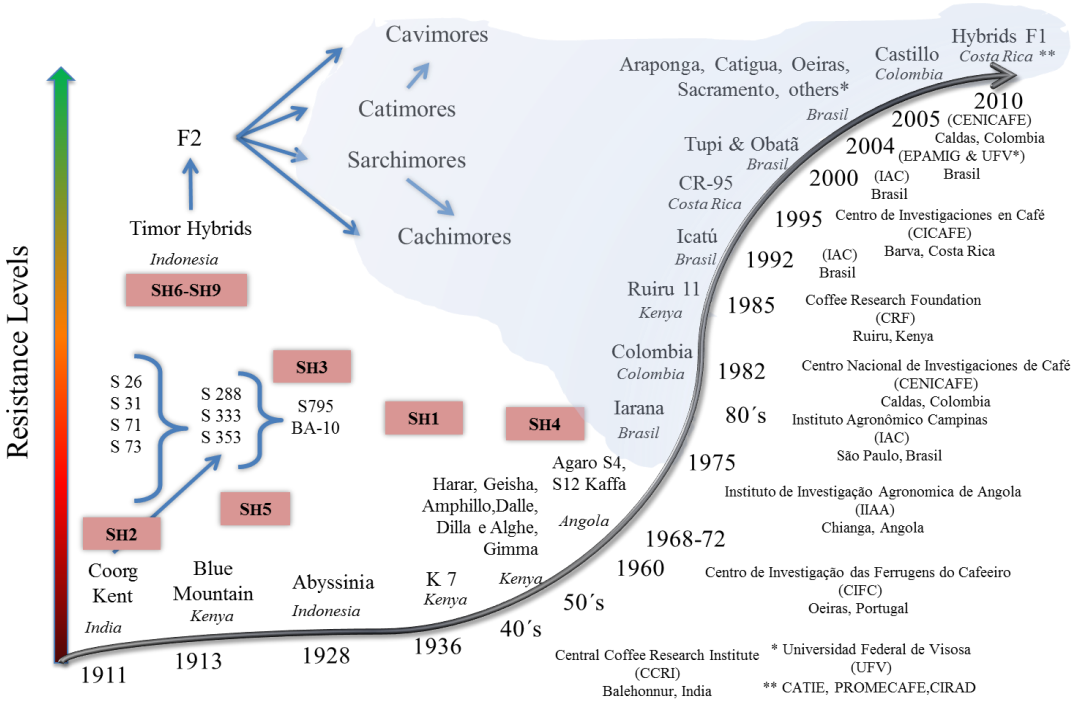


Figure 4. Schematic timeline of genetic resources released with some CLR resistance.

Depending on the nature of the host-pathogen interaction, the plants generally exhibit two types of disease resistance: *A)* host resistance and *B)* non-host resistance. Host resistance is related to a gene-for-gene interaction explained by Flor [41], where *R* dominant disease resistance gene products on the host recognize *avr* gene products or effectors of the pathogen, leading to a chain of molecular reactions from both organisms to defend or attack. The final goal of the host response is to eliminate the pathogen, or at least reduce the growth or colonization, so a hypersensitive response (HR) can be obtained to alert neighboring cells. This HR incomplete resistance normally leads to host defense responses in the cell and near the infection site, inducing programmed cell death

in the area to avoid the colonization of the pathogen [47-48]. The non-host resistance type of disease resistance exists when no specific *R* or *avr* product recognition exists, but both constitutive and inducible reactions prevent the specific host-pathogen interaction [49]. Host and non-host resistance are controlled by the genetics of both the host and the pathogen.

The phenotype is what characterizes the genetics of the resistance. According to the variability, a discrete class of distribution within the population of plants can indicate that fewer major genes are in charge of the resistance or susceptibility to the specific pathogen; therefore, this is related to a qualitative (vertical) mediated resistance, mostly based on host type resistance. However, a continuous distribution of resistance within a population of plants may indicate a higher quantity of genes with minor to moderate effect, related to quantitative (horizontal) resistance that can be associated to a non-host (non-specific) type of interaction [50]. Quantitative resistance can be expressed as an incomplete resistance, which is the host's ability to withstand high levels of the pathogen without a fitness cost [51]. Plant tolerance seems to be related to quantitative resistance in the host, but led by the pathogen's ability to suppress the plants immunity which results in a reduction of the disease, but not in its absence. To survive, the plants have to use both the specific and non-specific genes to defend against specific or non-specific pathogens, but also to properly regulate the immune system by diverse detection and signaling mechanisms [52].

At a molecular level, the immune system of a plant host and pathogen interaction is a dynamic signal and recognition crosstalk (Figure 5). Based on the host and non-host

resistance, different models can be used to explain the interaction using a basal resistance [53], zigzag [47], or quantitative and qualitative resistance [52] in the immunity system [54]. Briefly, all models involve first that the pathogen release microbe/pathogen/damage-associated molecular patterns (MAMPs, PAMPs, or DAMPs) during its penetration, which are recognized by transmembrane pattern recognition receptors (PRRs) in the host cell. The PRR receptors start a PAMP-triggered immunity (PTI), which can halt pathogen colonization by basal disease resistance (quantitative derived products) or specific *R* genes (qualitative) in the host. If the pathogen is able to detect the host barrier, it can further attempt to continue its invasion by the release of other effectors (avirulence products, *avr*) that will suppress PTI and induce an effector-triggered susceptibility (ETS). In order to recognize the specific *avr*, the host has to synthesize proteins with nucleotide binding sites and leucine rich repeat domains (NBS-LRR) among others. Those NBS-LRR proteins are coded by *R* genes and are involved in an effector-triggered immunity (ETI), which works as an accelerated and amplified PTI response, resulting in a hypersensitive response (HR) or cell death response at the infection site. Further HR and/or quantitative traits for resistance (QTR), such as broad spectrum, durable, or resistance related genes, are activated by hormones such as salicylic acid (SA) or cytokinin (Cyt), to enhance systemic acquired resistance (SAR). SAR prepares neighboring cells for a secondary attack and remains as a molecular alert.

Natural selection in an evolutionary perspective results in new *R* and *avr* specificities for vertical resistance, as well as PAMP's, PRR's, and QTRs for a more horizontal or basal resistance, so that ETI will continuously be diversified by

recognition, suppression, and modification processes. Thus, the ability of a host plant to survive over generations is dependent on the maintenance and generation of new variability resistance loci and signal reception.

The corresponding integrated model can explain the compatible interaction that occurs during several cycles of infection and recognition (C-G in Figure 5), leading to an explanation of why new CLR races can appear in coffee (qualitative resistance), and how different levels of resistance exist (quantitative resistance). However, some alternative hypotheses of why CLR is diversifying into new races may be related to mutations (most probable), sexual hybridization of the teliospores (not known yet), cryptosexuality, or horizontal gene transfer (HGT) that occurs in several oomycetes, such as *Phytophthora sp.* and *H. arabidopsidis* [55]. Another hypothesis may be related to epigenetics: inherited small RNA's seem to also serve as signals for internal and environmental conditions from parent to offspring or cell to cell. It is expected that those types of regulation mechanisms may also play a role in the aggressiveness and scaffolding of the pathogen over the coffee host on recurrent CLR life cycles [56].

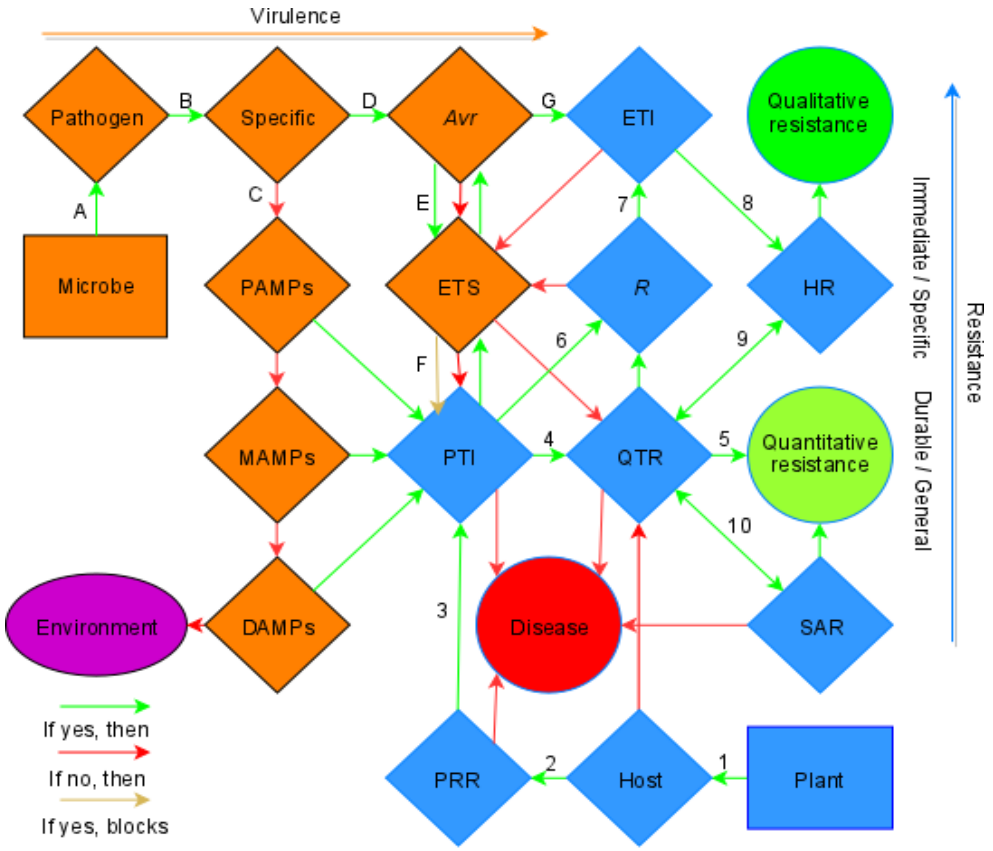


Figure 5. Integrated model explaining qualitative and quantitative resistance in the pathogen and host interaction.
The rectangles represent the microbe (pathogen) and the host (plant). The diamonds represent the conditional molecule or interaction according to the origin (orange colored related to the pathogen and blue for the plant). Arrows in green represent an affirmative condition, negative in red, and yellow for an effect that blocks the following condition. The circles represent the expected final response from the network. From left to right and top to bottom, the acronyms are as follows: PAMPs: pathogen-associated molecular patterns; MAMPs: microbial-associated molecular patterns; DAMPs: damage-associated molecular patterns; *avr*: avirulence genes; ETS: effector-triggered susceptibility; PTI: PAMP-triggered immunity; PRR: pattern recognition receptors; ETI: effector-triggered immunity; *R*: resistance gene; QTR: quantitative resistance genes involving broad spectrum and durable resistance genes; HR: hypersensitive response; SAR: systemic acquired resistance. A qualitative resistance may follow the path A, B, (C, F, E, D, G, or D, G) with 1-3, (6-8, or 4,9), while a quantitative resistance may be obtained by the same path in the pathogen, but 1-5 or 1,5 in the plant. Different networks may occur in the dynamic crosstalk between the pathogen and the host. Failure in points 2, 3, 4, or 10 may explain the host susceptibility.

Sequence comparison of plant NBS domains related to *R* resistance proteins revealed high conservation across a wide number of genera [57]. The explanation on how NBS domains diversified but retain functionality as pathogen induced signalers, may be the reason why even NBS domains can trigger defense responses by themselves, such as innate immunity and apoptosis responses in animals. The most common element of the NB domain is the Apaf-1 - R proteins - CED-4 (ARC) site, which functions as an ATP/GTP binding and hydrolyzing pocket that works as a transducer of ATPases and downstream signaling [58]. Two structural classes of *R* isolated proteins are known: the ones that include in their N-terminal domain a Toll and Interleukin-1Receptor (TIR) homology domain, and the others that have a coiled-coil (CC) or also called leuzine-zipper motif (non-TIR) [59]. Both TIR and CC domains indicate a role in recognition specificity and purifying selection. The variability of the N-terminus of the CC-NBS-LRR proteins appears to be one of the most significant sources of variation and probably specificity. On the other hand, in the C-terminus of the NBS domain, the LRR domain is also amino acid variable and under positive diversifying selection, suggesting a role in recognition specificity or protein to protein interaction [57, 60]. Furthermore, *R* proteins have been found to be clustered to multiprotein resistance complexes, which increase the potential of recognition and signaling against pathogen invasion [61].

Some NBS-like *R* genes have been identified in *Coffea* [57, 62]. Noir, et al. [57] reported nine distinct classes of resistance gene analogs (RGA's), all belonging to the CC-NBS type and with a sequence length varying from 480 to 600 bp. High similarities (35-55%) of the *Coffea* CC-NBS *R* genes to those in *Arabidopsis*, tomato, and rice has

been reported. The accumulation of mutations between the 9 RGA families is higher between them than within them, suggesting that all groups had a different evolution divergence. More recently, Ribas, et al. [62] reported that one of the most durable loci that controls CLR (S_H3), may correspond to a complex multi-gene cluster of CC-NBS coding genes, validating the idea that major resistance and stress related genes may be coordinated to increase the resistance of coffee to CLR, and why more complex races of CLR may appear.

Transcription factors (TF) and micro RNAs (miRNAs) are involved in the expression and regulation of several genes related to stress and many other networks [63-64]. Some TF's belonging to the highly conserved WRKYGQK motif and a C_2H_2 or C_2HC zinc finger (WRKY) family proteins, have been shown to be involved in the activation of defense response genes under stress [38]. For example, some hypersensitive response (HR) plant immune defense system pathways that regulate the accumulation of phenolic compounds and callose at infection sites are part of the responses mediated by WRKY with others such as non-race specific disease resistance 1 (*NDR1*) genes [65]. Gossypol, capsidiol, berberine, camalexin, and caffeic acid are other examples of the phytochemicals synthesized by secondary metabolic pathways regulated by WRKY TF's, which function as disease antagonists and in signaling mutualistic interactions with other insects [66-67]. In a general perspective, WRKY TF's can regulate the production of metabolic pathways related to phenylpropanoids, indol alkaloids, and mono-, di-, tri-, sequi-, and polyterpenes. Those metabolites confer functional and structural protection against a variety of abiotic stresses, including wounding, drought, salt, heat, cold, and

osmotic pressure. Moreover, the regulation of those metabolites has been shown to protect against biotic stresses as well.

The activation of WRKY TF's is also regulated by the accumulation of hormones such as jasmonic acid (JA) and salicylic acid (SA) [68]. The association of WRKY mediated responses related to JA induction is predominant on defense mechanisms against necrotrophic pathogens that kill host cells for nutrition and reproduction, whereas SA induced responses are more related to defense against biotrophes that require keeping the host alive in order to complete their life cycle, such as nematodes and rusts like CLR. The signaling network between JA and SA is also regulated by other hormones, for example gibberellic acid (GA) mediated by DELLA proteins[69]; abscisic acid (ABA), and ethylene in antagonistic SA responses; and auxin and cytokinin suppressing either SA or JA in order to prioritize plant growth. Hormone regulation of SA and JA signaling pathways are networking in coordination with the transcriptional and post-transcriptional regulation and TF's [68]. The networking and regulation of defense genes likely represents a core component in the coordination of multiple biological processes to favor defense over development [48, 70].

The control of the activation of defense responses seems to be a more durable strategy for controlling coffee resistance against CLR. In coffee, more than 49 WRKY proteins have been associated with defense response pathways induced by JA, SA, or CLR [66, 71]. More specifically, Caarls, et al. [68] reported that WRKY homologous genes, such as *CaWRKY11*, *CaWRKY12*, *CaWRKY6*, and *CaWRKY13/14* are activated by pathogens (rust and nematodes) and hormonal treatments (SA and/or JA), but not by

wounding, suggesting the idea that specific host-pathogen proteins may exist in order to jump from ETS to ETI responses. As an example of the gene-for-gene concept explained by Flor [41], Gil Azinheira, et al. [72] showed that non-host interaction of CLR race II in *Arabidopsis thaliana* produces a strong activation of WRKY and a pathogenesis related gene *PR1b* (an SA responsive gene) after 18 hours of inoculation (hai), when the penetration of hypha of CLR is probably detected. Even when the fungus successfully entered the stomata, the haustoria failed to develop, suggesting that specific *H. vastatrix* haustoria-formation signals may be related to the detection of *R* genes from the coffee host after ETS response, which later triggers the ETI response. Therefore better resistance of the coffee plant may be achieved by non-host resistance genes.

It is now recognized that the expression of non-host resistance genes is more likely durable and effective than complete resistance carried by race specific genes [73]. In wheat, for example, the use of several partial pathogen resistance genes such as *Lr34*, *Lr46*, *Lr67*, and *Sr2* combined with other minor genes, has led to a more durable approach to rust resistance [74]. Thus, inferring from other crops, the availability and expression of several metabolic and resistance genes can be a more sustainable alternative for CLR control, rather than preserving the dominant resistant genes alone.

In nearly all crops, genetic resistance to pests and diseases is desirable, and the most economically reliable method of resistance. However, a commercial variety also needs to have competitive quality and yield in addition to disease resistance. Those characteristics are part of the many reasons why the breeding process is extended and difficult. The market for Arabica only accepts specific varieties with good quality

attributes, and the growers prefer a high yielding and resistant plant. Reaching all those characteristics is the big challenge.

From a breeding perspective, it is important to establish a pyramid of genes for pathogen resistance to ensure durable resistance to many pests and diseases, especially focusing on the development of better non-host resistance, without losing quality attributes [73]. Some interesting quantitative trait loci (QTL) markers have been found for some diseases such as the S_H3 locus on chromosome 1 for CLR resistance, S_H5 that gave resistance to CLR race II (*avr* 5), *Ck-1* for the coffee berry disease (CBD) resistance on chromosome 11, and *Mex*-1 for resistance to *Meloidogyne exigua* on chromosome 4; however, accumulating genes in one genotype is difficult using traditional methods [6, 73]. If sets of resistance conferring genes can be included with other known quality improvement genes, the breeding of coffee can be dramatically expedited.

1.1.5 Quality

Achieving an excellent cup of coffee is still an art due to the complexity of factors involved. Quality depends on the genetic origin (species and genotypes), non-genetic sources of variation (terroir, ripening time, post-harvest, etc.), and consumer preference, which all converge into the bean chemistry and transformation [75]. The chemical composition of the bean can be affected by any of the previously mentioned factors; however, the most impactful quality factor at pre-harvest is the plant's genetics, while at the post-harvest stage, the roasting process has a significant impact on quality. Both of these factors are related to the ripeness of the fruit.

Bean ripening time determines the availability of various chemical compounds, but the type and quantity of these depends on the first stages of fruit development (Figure 6) [76]. Phenologically, fruit development can be divided into immature and mature stages, reaching the maximum endosperm growth and ripening at the cherry stage between 210 and 240 days after pollination (DAP). During the first immature stages of development (1-3 in Figure 6), the fruit behaves more like an active leaf with a high level of cell division and expansion, synthesizing carbohydrates and lipids that are going to serve as sources of energy and storage. Gene expression during the first stages is related to the biosynthesis of pectin / cellulose, cell wall polysaccharides (CWP), lipids such as the ones derived from the diterpenes cafestol and kahweol, and chlorogenic acids (CGA) as precursor of lignin. The biosynthesis of alkaloids such as trigonelline and caffeine derived from the phenylpropanoid pathway are also involved during the first growth stages [77].

At the yellowish-green stage, the endosperm occupies ~60% of the fresh weight of the fruit [78], and myo-inositol, sucrose, and/or raffinose family oligosaccharides accumulate as osmoregulators [79] or storage for later embryo germination [77]. α-galactose and other sucrose transporting enzymes continue the CWP biosynthesis and mobilization to the endosperm. The same precursors coordinated by phenylalanine ammonia-lyase (PAL), promote the generation of more lignin, caffeine, flavonoids, and help the accumulation process of carbohydrates and oils [78]. Fatty acid and triacylglycerol (TAG) biosynthesis is also increased by pyruvate dehydrogenase (PDH), ketoacyl-ACP synthase (KASII), and acyl-ACP thioesterase (FATb), which respectively synthetize the

acetyl-CoA (oxidase from pyruvates obtained from glycolysis), 18:0-ACP (precursor of the linoleic acid 18:2), and trigger the entry of 16:0 (palmitic acid) to the acyl-CoA into the endoplasmic reticulum which enlarge the TAG's [77].

The maturation stage (3-5 in Figure 6) of fruit formally starts with ethylene production from the aminocyclopropanecarboxylate (ACC) oxidase (ACO), which consumes S-adenosyl methionine (SAM) derived from the ACC synthase activity [80]. The ethylene signal is received by ethylene receptors (ETR) to downstream activate/deactivate genes such as sucrose synthases (SUS) and (1‑4)‑beta‑mannan endohydrolase (MAN-B) for sugar accumulation during the ripening, and/or other transcription factors (TF's) which modulate perisperm thickening and anthocyanin (or luteolin) accumulation resulting in change of color [80-82]. Once the highest sugar concentration is achieved, the pericarp is softened by the pectin methylesterases [83], while ethylene-responsive factor (ERF) and alcohol acyltransferase increase their expression [82]. The alcohol acyltransferase is related to the synthesis of ester types of volatile compounds, derived from the acyl-CoA, which are related to plant defense and plant-to-plant signaling [84]. During this last stage it is expected that the fatty acids, amino acids, proteins, and carbohydrates that have accumulated in the endosperm are going to be available as precursors of volatile compounds, together with the trigonelline and CGA's synthesized during the immature stage [85]. After ripening, the isocitrate lyase, associated to germination processes, is the marker of the beginning fruit senescence and drying process [76].

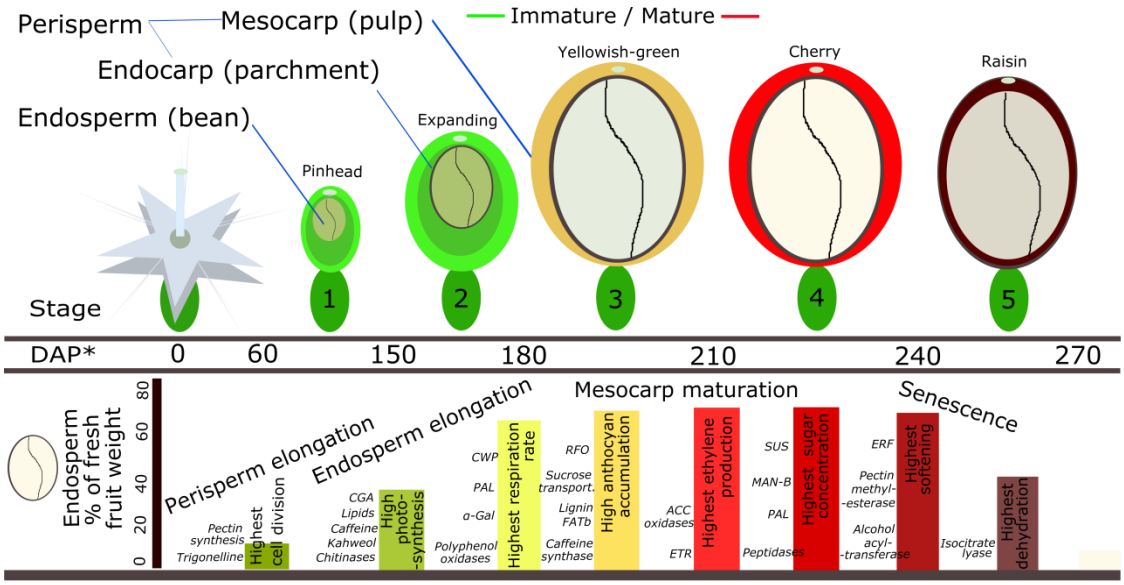


Figure 6. Schematic representation of the tissues, common compounds, and genes involved during the fruit development stages.
Details in [77-78, 82-83].

During the post-harvest period, ripened fresh fruits have to be collected and later mill processed to obtain the beans. Two methods can be implemented: wet (or washed), where the cherries are depulped and later fermented for 16-36 hours in water; and dried (also called natural), were fresh fruits are dried for 10-20 days and later depulped. De Bruyn, et al. [86] showed that both methods significantly influenced the microbial communities and hence, the composition of the final green coffee beans. The choice of post-harvest wet or dry processing therefore may explain some differences in the final beverage quality. For example, dry processing generally enhances bitterness in the final liquor, while washed coffees have a better perception overall [87].

The chemical composition of the green bean is transformed through heat and non-enzymatic reactions during the roasting process [88]. The roasting process involves

reactions such as Maillard reactions (browning reactions between sugars and amino acids), Strecker degradations (amino acids reacting with carbonyl grouped molecules to form aldehydes and ketones), and caramelization (transformation of complex sugars into more simple sugars) [15, 89-94]. According to the profile of temperature and length of roasting, various coffee flavors may be changed due to the formation of chemicals like nicotinic acids, chlorogenic acids (CGAs), and volatile compounds [5, 87, 95]. For example, nicotinic acid (derived from trigonelline) and pH, were found to increase with time and temperature [95], while caffeine concentration deceased under the same conditions [96]. Increasing the temperature and time to give a light (roasted to just before first crack), medium (traditional roast in the eastern United States, roasted to the end of the first crack), city (medium brown, a typical roast throughout the United States), or French (dark roast obtained at the end of the second crack) type of roasting, linearly increased the content of the volatiles pyridine, catechol, and ɣ-butyrolactone, while decreasing 2,5-dimethyl-4-hydroxy-3(2H)-furanone, 5-hydroxymethylfurfural, and 3,5-dihydroxy-2-methyl-4H-pyran-4-one [96]. Similarly, Schenker, et al. [97] showed that the use of higher temperature increased 2,3 butanedione, 2-ethenyl-5-methyl pyrazine, 4-4-hydroxy-2,5-dimethyl-3(2H)-furanone, 2-2-hydroxy-3-methyl-2-cyclopenten-1-one, linalool, kahweofuran, 2-methyl butanal, 3-methyl-2-buten-1-thiol, 2,3-pentanedione, and propyl pyrazine, with aromas corresponding to buttery, musty-burnt, caramel, spices, flowers, coffee, malt, sulfur-like, butter, and potato, respectively. Redgwell and Fischer [98] reported that increasing temperature linearly increased polysaccharide degradation and solubility, especially of arabinose, galactose, mannose, rhamnose, pectin, and glucuronic

acid, and showed it to be genotype dependent. The creation of new chemical compounds and combinations from the green beans results in different sensory profiles.

Once the roasting process has transformed the green bean chemistry, the important flavor and aroma attributes can be perceived and the beverage quality profile determined. For example, the flavor profile can be positively affected by sucrose, trigonelline, and lipids, which are referenced as the major non-volatile biochemical compounds contributing to good quality of the roasted bean. In contrast, caffeine, some CGAs, and several compounds in the groups of alcohols, aldehydes, hydrocarbons, and ketones are associated with bad coffee quality [99-100]. The aroma is generated from isoprenoids, phenylpropanoids, amino acids, and fatty acids during the Maillard reaction and is one of the most important and complex factors involved in coffee quality [101-104]. For example, aroma perceptions of meaty, roasty, potato, catty, earthy, spicy, and honey, are assumed to be related to 2-methyl-3-furanthiol, 2-furfurylthiol, methional, 3-mercpto-3-methyl-butyl formate, 3-isopropyl-2-methoxypyrazine, 4-ethylguaiacol, and (*E*)-β-damascenone, respectively [88]. Although there are more than 1000 volatile compounds produced during coffee roasting, the molecules responsible for good quality coffee are still unclear.

One major concept that has to be addressed for quality is that the chemical composition of the bean didn't evolve for human pleasure. All sugars, proteins, lipids, and secondary metabolites are intended as nutrient supplies for the embryo and survival (though survival might have been improved by producing compounds of interest to other organisms). For example, endosperm cell walls are rich in insoluble polysaccharides

(>50% w/w), mainly composed of cellulose and hemicelluloses, and soluble carbohydrates (>40% w/w), such as the monosaccharides (fructose, glucose, galactose and arabinose) and oligosaccharides (>90% sucrose) [105]. Accompanied proteins (9-12% w/w), lipids such as triacylglycerols, sterols, tocopherols, diterpenes of the kaurene family (8-18% w/w), and other minerals (3-5% w/w) inside the endosperm cytoplasm, are also important to embryo development [105-106]. Some phenolic compounds such as esters of hydroxycinnamic and quinic acid, collectively known as chlorogenic acids (CGA), represent up to 12% of solids. Secondary metabolites of plants such as tannins, lignans, and anthocyanins, are in lower amounts in the bean, and are generally involved in defense against ultraviolet radiation or aggression by pathogens [107].

The main alkaloid synthesized in coffee is caffeine, which varies from almost no content in the Mascaracoffea species up to 2.7% in *Coffea canephora* from the West and Central African species. Caffeine is presumably an adaptive defense repellent against increased rates of insect predation like the coffee berry borer (CBB, *Hypothenemus hampei*) [14]. It is assumed that roasted coffee is composed of 38–42% w/w carbohydrates, around 23% melanoidins (high molecular weight compounds of unknown structure), 11–17% lipids, ~10% proteins, 4.5–4.7% minerals, 2.7-3.1% CGA's, 2.4% aliphatic acids, 1.3-2.4% caffeine,~40 known volatile compounds that contribute to the aroma, and other active compounds such as serotonin and its precursors L-tryptophan and 5-hydroxytryptophan [105]. However, all previous chemical compounds are physiologically involved in the organism survival. The dynamic and complex profiles of the resulting

brewed beverage are then a multifactorial history of the plant's response in order to preserve the species.

Since the precise molecules involved in a good cup of coffee are unknown, it is also unknown if these specific molecules are necessarily consistent to a sensorial perception or genetic origin. This is likely because few precise research tools are designed to discover or verify the interrelationship of these factors. Quality is determined by a complex mixture of somatosensory inputs and social influence. The senses, such as olfaction, gustation, oral, and nasal, interact with cultural preference and marketing, thus giving different notions of what quality is [87-88]. Tasting or "cupping" by professionals is one method traditionally used to determine the organoleptic characteristics of a coffee stock. During the sensory analysis, the panels of judges utilize their training and experience to transform their perception into numbers and categories. The results are statistically analyzed to later interpret the results. Uniformity and repeatability of a protocol is required.

One protocol is the conventional Specialty Coffee Association of America (SCAA), used to determine the fragrance, flavor, acidity, body, balance, sweetness, uniformity, cleanliness, and scoring of each sample on a scale from 0 to 10 [88]. Another recently developed tool is the WCR Sensory Lexicon, that uses an intensity scale to describe quantitatively the qualities expected to differentiate coffee attributes in a cup [108]. However, since all the information is provided by humans, relevant evidence of the chemical composition related to those perceptions is missing.

During the last decades, several strategies for objectively assessing a specific score of a coffee have been developed. Physical analysis by near infrared reflectance spectroscopy (NIRS) or nuclear magnetic resonance (NMR), and chemical analysis, such as high-performance liquid chromatography (HLPC) and solid phase micro extraction (SPME) followed by gas chromatography mass spectrophotometry (GC-MS), are examples of techniques used to describe and determine specific compounds in coffee [89, 102, 109-111]. The use of these types of technologies, increase the knowledge about the biochemical compounds that can be related to coffee quality.

1.1.6 Molecular genetic tools: transcriptome information

Molecular tools have been used on coffee to understand the origin and genetic resources since the late 1980's and these tools have continued to evolve in coffee research [112]. The first application of molecular markers was reported in 1982, studying the DNA variation of the chloroplast and mitochondria of *Coffea* [113]. Later in 1993, randomly amplified polymorphic DNA (RAPD) markers were used to screen inter- and intra-specific *Coffea* variability [114]. Three years later, in 1996, DNA internal transcribed spacer (ITS2) regions and chloroplasts were targeted to find phylogenetic relationships of *Coffea* species, and validate the amphidiploid origin of *Coffea arabica* [115-116]. Amplified fragment length polymorphisms (AFLPs) and simple sequence repeats (SSRs) were used to estimate genetic diversity and introgression in *C. arabica*, [57, 112, 115, 117], diversity , but also to find resistance gene analogs (RGA) and related resistance genes [118]. The genetic linkage map of *C. canephora* was completed by Lashermes, et al. [119].

Inter-simple sequence repeat (ISSR) markers were used to find diversity in wild *Coffea* populations [120], expressed sequence tags (EST's) to find more SSR's for comparative genomic studies [121], and bacterial artificial chromosome (BAC) libraries for functional and comparative genomics [122-123]. The chloroplast genome of *C. arabica* was released by Samson, et al. [124] in 2007, which was visualized as a better resource for genetic transformation. With ribonucleic acid (RNA) technology, transcription studies started to unravel regulation mechanisms of transposable elements (TE) [125], and contrasting gene expression levels under abiotic stress [126]. In 2014, the draft genome of *Coffea canephora* was released, which was enriched with the creation of web repositories such as the Brazilian coffee genome project, the CoffeebEST [127], and CoffeeHub [128]. The most recent molecular tool is a 26K SNP (single nucleotide polymorphism) chip for *C. canephora* [129], which may be used to facilitate GWAS and genomic predictions, which may help future breeding programs.

The use of molecular markers allows rapid and efficient inference of the presence of more than one desirable gene or locus. For example, Zeng, et al. [130] justified the use of gene pyramiding for breeding high-yield and superior-quality rice varieties. They made three-way crosses between well characterized elite varieties with major-known gene repertoires for each quantitative trait, plus minor-unknown genes, using QTL markers. The reduction of time and increased accuracy resulted in new improved varieties during a five year breeding program. In maize, durable resistance has been widely used in breeding programs for quantitative traits such as the resistance to necrotrophic diseases [131]. The discovery of quantitative resistance genes has been made possible through the

use of multi-parent advanced generation intercross (MAGIC) and nested association mapping (NAM) populations, which conferred enough statistical power to find minor QTL's which were later converted to markers for validation studies [132]. Thus, when the objective is gene pyramiding with high quality genes and/or durable resistance, molecular tools may provide several advantages to plant breeders [44].

The possibility of predicting phenotypic features based on genetics is relevant towards long-term agronomic improvement of coffee and many other crops [133-137]. For example, most of the important agronomic characteristics of crops are quantitative traits, such as yield, quality, disease and pest resistance, and adaptation. The genetic variations for those quantitative traits are generally described as additive effects, biasing from the expected multifactorial dynamic of genetic networks and minor genes [138]. The result is that breeding studies using quantitative trait locus (QTL) markers and genome-wide association study (GWAS) information, are not fully, accurately, or usefully predicting the relevant chromosome regions or genes. The problems with detection of the relevant chromosome regions are related to a lack of power to detect loci contributing small effects, differences in genetic background, epistasis, and repeatability in other population structures [137]. Nevertheless, much of the plant molecular technology used today is based on measurement of DNA as opposed to RNA or proteins, because it is less expensive to process, can be obtained from almost any tissue, is more stable, and it is also directly related to the genomic information regardless of environmental conditions. However, what makes the DNA molecule desirable for molecular studies is also the limitation in discovering genes that represent the phenotypes under different conditions. Even when it

is assumed that the DNA molecule is stable enough to conserve the integrity within the organism, this doesn't explain how DNA is related a phenotype.

The logical basis of how DNA represents a phenotype, requires several steps conceptualized in the central dogma of molecular biology, which has as one of the components, the molecule of RNA [139]. Three main processes are distinguished in the central dogma of molecular biology: DNA replication, RNA transcription, and protein translation, each having their own diverse regulation and control mechanisms (Figure 7) [139-141]. Even when the regulation processes are complex and still not totally understood in all contexts, the most important product of DNA transcription (the RNA transcriptome), is one of the best approaches to find the relationship between the environmental stimulation, cell sensing, and the organism's molecular and phenotypic response.

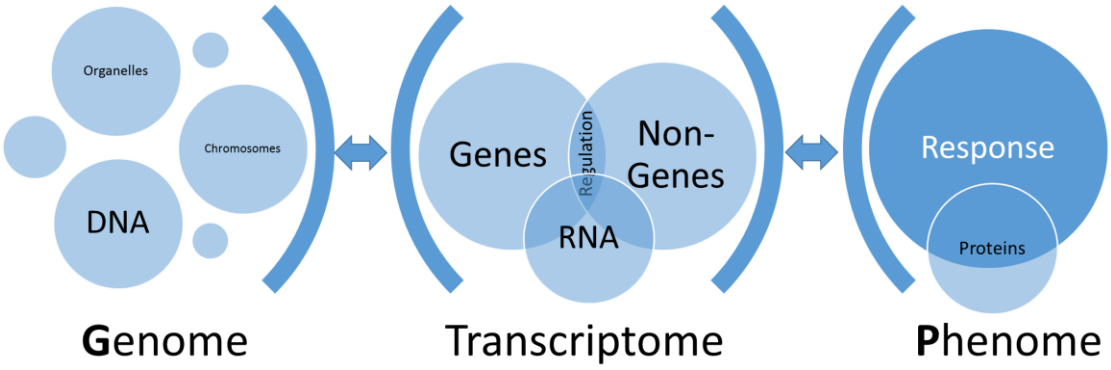


Figure 7. Modified central dogma of molecular biology concept representation.

When many different samples are examined, the information obtained from the RNA sequencing process gives a gene expression "atlas", which enables the

understanding of various environmental perturbations and how the plant responds at the genetic level [38, 142-143]. Moreover, the transcriptome is related to the molecular characterization of the organism in a specific moment and stimulus, but also reveals part of the genome [144-145]. In genetic improvement research, the transcriptome is valuable if further molecular, computational, and comparative analysis can be made to predict the RNA sequences' biological or regulatory function related to the observed phenotype, as a potential source of variation that can be exploited in breeding material.

Currently, high throughput next generation sequencing (NGS) techniques, such as mRNA sequencing (RNA-seq), have been widely used in studies of plant transcriptomes in several crops. With the utilization of NGS tools, the identification of mutations and new alleles responsible for major genetic effect variants are being accelerated, even faster than microarray technology. NGS tools and expression quantitative trait loci (eQTL) markers have higher potential to provide refined information at the isoform, transcript, and allelic expression levels [146]. A combination of gene expression studies by eQTL or quantitative real time polymerase chain reaction (qRT-PCR), with traditional linkage studies on structured populations, such as GWAS, may help to predict the expression level and regulation in polymorphic regions where quantitative traits may be associated [147]. For example, transcriptome and GWAS data was used to predict 14 candidate genes related to yield in *Brassica napus*, which remain to be validated. Because yield is a result of a complex interaction between genotype and environment during the course of crop growth and development, QTL data alone won't be informative enough to reveal the chromosome regions and loci involved in the

expression of the phenotype, especially if the QTL is highly environmentally dependent. Therefore, combining differentially expressed gene information with significant loci affecting yield traits increases the likelihood of selecting candidate genes with predicted functionality and relevance.

Coffee displays among the most conservative chromosomal gene order and functional synteny among asterid angiosperms, with a one-to-one chromosome correspondence with grapevine and one-to-three with tomato [19, 99, 134]. Both subgenomes in *C. arabica* (E^a and C^a), show 95% similarity to one another due to coffee's low sequence divergence, almost perfect gene synteny, and high number of shared transposable elements (25%) [20]. Since the recent speciation process occurred 10-450 thousand years ago, the regulation and expression of the genes involved in all molecular, cellular, and biological processes may still be undergoing selection processes in *C. arabica* [125]. Therefore, it is likely that duplicated homologous *trans*- regulatory factors (genes which may regulate the expression of distant genes), interact and control the transcription for both subgenomes, increasing the combinations of genotype responses to stimuli and plasticity response [126].

The gene expression of *C. arabica* is regulated distinctly from both progenitors´s genomes. The genome size of *C. arabica* var Geisha (UCG-17) recently released by Medrano, et al. [148], was estimated to be 1.11 Gbp, with an estimated number of 70,830 genes. The *C. canephora* genome size is 710 Mbp, and is predicted to have 25,574 protein coding genes [134]. Comparative transcriptome analysis showed that *C. eugenioides* may have high similarity to other *Coffea* species; however, it has a

differential expression of genes related to small molecule binding, transferase activity in sugar synthesis, chitinases, and others [19]. Similar to *C. eugenioides*, *C. arabica* produces and accumulates more sucrose during grain development compared to *C. canephora* [91]. Another study [149] revealed that the homeostasis of membrane fluidity and ROS (reactive oxygen species) genes under different thermal growing conditions were increased in *C. arabica* when compared to genotypes with just one of the parental subgenomes, suggesting that the interaction of both genomes in *C. arabica* may be improving phenotypic homeostasis under unsuited environments compared with the progenitors.

RNA-seq data can help to understand the importance of the metabolic networks and the response of the coffee plant to stress. For example, the most updated transcriptome analysis in coffee leaves and fruits made by Ivamoto, et al. [79], discovered differential expression of galactose metabolism and raffinose biosynthesis-related (RFO) genes under abiotic stress. The same authors validated the candidate genes by qRT-PCR and showed that some of those RFO genes were up-regulated under water deficit, highly salinity, cold, or heat stress conditions, acting as osmoregulators in leaves and some development stages of fruit formation. Physiological and molecular studies will open new horizons in the knowledge of how breeding can improve disease and pest resistance, climate change resilience, and improve quality related traits. The development of alternative methods based on the knowledge of the gene expression on various context linked to the phenotype, can help to increase the resolution efficiency compared to using GWAS alone.

Future challenges in agriculture under a changing climate may be devastating if the comprehension of the physiology, genetics, and environment remains unconnected. The objective of the present research was to identify candidate genes expressed in the coffee plant, during stressful conditions exacerbated by rust disease and yield level after manually thinning or no-thinning, and to understand how these interactions affect cup quality and plant performance. The expression profiles from the sequences analyzed will allow the identification of metabolic pathways and associated candidate markers, so that later studies can deliver more information about plant health conditions and determine the best technologies to mitigate those impacts.

CHAPTER II

RUST AND THINNIG MANAGEMENT EFFECT ON CUP QUALITY AND PLANT PERFORMANCE FOR TWO CULTIVARS OF *COFFEA ARABICA* L. [1]

Beverage quality is a complex attribute of coffee (*Coffea arabica* L.). Genotype (G), environment (E), management (M), post-harvest processing, and roasting are all involved. However, little is known about how G x M interactions influence beverage quality. We investigated how yield and coffee leaf rust disease (CLR) (*Hemileia vastatrix* Berk. et Br.) management affect cup quality and plant performance, in two coffee cultivars. Sensory and chemical analysis revealed that 10 of 70 attributes and 18 of 154 chemical volatile compounds were significantly affected by G and M. Remarkably, acetaminophen was found for the first time in roasted coffee and in higher concentrations under more stressful conditions. A principal component analysis described 87% of the variation in quality and plant overall performance. This study is a first step in understanding the complexity of the physiological, metabolic, and molecular changes in coffee production, which will be useful for the improvement of coffee cultivars.

---

[1] Adapted with permission from Fabián Echeverría-Beirute, Seth C. Murray, Patricia Klein, et al. 2017. Rust and Thinning Management Effect on Cup Quality and Plant Performance for Two Cultivars of *Coffea arabica* L. Journal of Agricultural and Food Chemistry.

2.1 Introduction

Achieving an excellent cup of coffee is still an art even beyond an individual's taste preference. Consider the entire journey of the bean from the plant to the cup. The plant requires optimal development for flowering and pollination; the fruit has to grow and mature under favorable conditions, as well as undergo harvest at the optimal stage. The fruit must be suitably transported to a milling facility, depulped, dried, and the parchment depleted. The resulting beans have to be transported to a roasting facility and then to the end user without suffering any degradation. Finally, there are a myriad of alternative techniques to grind and brew the coffee. All the cumulative effects of each beans journey may create different results both additively and in interactions with other factors. Each factor consequently, is a starting point for scientific exploration of what determines quality in a cup of coffee. However, scientific investigation requires that many of these variables be held constant, repeatable screening methodologies be developed, and sample sizes be large enough for detecting meaningful differences in variation.

2.1.1 Major management factors

The phenotype of coffee is a result of the interaction of genotype (G), management (M), and environment (E) [88, 150]. The environment includes the latitude, altitude, climate, and physical, chemical, and biological soil properties, which are uncontrollable factors that condition the agroecosystem. The management or agricultural

practices, can modulate the interaction of the plant and the environment through shade incorporation, thinning (reduction of berries in the tree), irrigation, fertilization, and pest/disease control; however, if the plant's genotype is not adapted to the specific environment, management practices will not be enough for adequate coffee production or quality.

From the production perspective, the environment and management are the main components that can modify the yield and quality potential [151]. Specialty coffee of *Coffea arabica* L. cultivars are generally produced under cooler climatic conditions (18-21ºC), exhibiting more acidity, better aroma, and fewer flavor defects than those produced in warmer regions [111, 152]. Field management practices can also increase sensory quality of the beverage. The use of shade trees increases sensory perception at the expense of a 15 to 25% reduction in yield [75]. Reducing berries on the tree through manual fruit thinning, has been hypothesized as a way of reducing biennial ripening stress on fully loaded trees, improving tolerance to pest and diseases, as well as inducing earlier and more complete fruit maturation, which is likely to enhance coffee quality; however it is not a common practice at many coffee farms [75, 153]. The impact on final coffee quality of major agronomic practices such as the addition of soil amendments, fertilizers, and chemicals for pest and disease control, are not clearly understood; however, are required for the compensation of nutrient fruit uptake, metabolism, structure, and physiological health maintenance for the plant's growth, production, and adequate response under stress conditions [88].

One of the major problems affecting coffee production worldwide is the challenge associated with the control of coffee leaf rust (CLR, caused by *Hemileia vastatrix* Berk. et Br.), which reduces photosynthetic leaf area and yield from 30 to 80% [4, 28, 87]. Combining coffee genetics with management practices is recommended for integrated control of the disease [26, 154-155]. Most cultivated coffee areas around the world are currently growing industry-accepted cultivars that are CLR susceptible. Therefore, management practices are primarily focused on disease prevention, and include activities such as spraying protectant coppers or systemic triazole or strobilurin-derived fungicides.

To understand the interactions between management practices and genotype on cup quality, requires a comparatively large number of samples and conditions be objectively evaluated. To date, there have been few experimental designs where multiple factors have been studied synergistically to understand the physiological and chemical changes of the green bean, which could have an effect on cup quality [75, 105, 111]. Furthermore, to our knowledge there has not been a report on how rust management methods affect downstream cup quality.

2.1.2 Importance of roasting on coffee quality

Roasting is one of the key steps towards the determination and description of the quality of coffee caused by the interaction and transformation of chemical compounds in the green bean [156]. Roasting involves thermo-induced non-enzymatic reactions between nitrogen containing molecules (e.g., free amino acids, peptides, polypeptides, and

proteins) simple and complex carbohydrates, lipids, and secondary metabolites (e.g., caffeine, trigonelline, chlorogenic acids, etc.) modified by three processes. The first process involves the Maillard reaction (browning reactions between sugars and amino acids), the second Strecker degradations (amino acids reacting with carbonyl grouped molecules to form aldehydes and ketones), and the third process by the caramelization (transformation of complex sugars into more simple sugars) [88, 110]. These and other physical and chemical changes generate most of the aroma and flavor related profiles in roasted coffee [85, 88, 157]. The complexity, diversity, and quantities of chemical compounds involved in beverage quality are still an active area of research.

2.1.3 Evaluation methods of coffee quality

Tasting or cupping is one of the most important methods used to evaluate the coffee quality. Because the perception of quality is a complex mixture of inputs of olfaction, gustation, preference and social-market interaction [87-88], tasting by professionals is used to determine the organoleptic characteristics. One cupping method used to evaluate the quality of the infusion is the Specialty Coffee Association of America (SCAA) protocol, which determines the fragrance, flavor, acidity, body, balance, sweetness, uniformity, cleanliness and scoring of each sample on a scale from 0.00 to 10.00 [158]. Another recent tool is the World Coffee Research (WCR) Sensory Lexicon, which uses an intensity scale to describe the qualities expected to quantitatively differentiate coffee attributes in a cup [159]. In both protocols, qualified individuals taste the beverage samples and provide information on their organoleptic characteristics

thereby giving room for human error and rendering the evaluation process subjective. Furthermore, it is challenging to make recommendations to coffee breeders and managers based on these evaluations [6]. For these reasons, it is useful to have inexpensive, rapid, and objectively measured chemical composition data to support taste panel perceptions.

Several technologies for objectively measuring specific chemical compounds of coffee were developed over the last few decades. Techniques used to describe and determine specific compounds in coffee (as well other crops such as maize, sorghum, and cotton) include, quantitative physical analysis by near infrared reflectance spectroscopy (NIRS) or nuclear magnetic resonance (NMR) , chemical analysis such as high-performance liquid chromatography (HLPC), and solid phase microextraction (SPME) followed by gas chromatography mass spectrophotometry (GC-MS) [102, 111]. Integrating human perception along with chemical information will allow for the evaluation of quality on a scientific basis, which will help to further research and development.

The aims of this study were to (1) measure G x M responses of two susceptible cultivars to CLR and yield control on plant overall performance and coffee quality, and (2) to compare quality as evaluated by cupping to objective chemical methods. This work will begin to elucidate the biochemical basis of intricate sensorial phenotypes of coffee beans produced under different management conditions and to establish a model that explains the interactions between plant overall performance and cup quality.

2.2 Materials and Methods

2.2.1 Experimental design

The experiment was located in the commercial coffee farm Hacienda Aquiares, located in Turrialba, Cartago province, Costa Rica, on the coordinates N 10º03´513 and W 84º13´583, at 1110 m.a.s.l, with an average temperature oscillating between 20ºC at night and 25ºC during the day, and 2200 mm annual precipitation. Two coffee (*Coffea arabica* L.) plots of mature plants were selected. One plot was planted with an inbred (Catuai vermelho IAC 144, F_8 originating from 'Caturra' x 'Mundo Novo', registration Nº 02929-SNPC, Brazil) susceptible to predominant races of rust, and the other plot with a hybrid (H3, F_1 of 'Caturra' x 'Ethiopian 531') with slight tolerance to the predominant races of rust. The experiment was performed using a split-split-split plot complete block design, where the main split was genotype, the second split was leaf rust control with four replicates within each block, and the third split was thinning. The second split was subdivided into two blocks: block one, which had fungicide application (rust control) with alternate sprays of cyproconazole and epoxiconazole [500 ml/hectare (Ha)] as active ingredients, and block two, without fungicide application (no rust control; Figure 8). The third split with the fruit yield treatment, was subdivided into complete fruit yield (0% thinning) and half fruit yield (50% thinning). Sixty days after the last flowering, the fruits were thinned by reducing 50% of the fruits on each branch of the plant by alternative harvesting of the nodes by hand. Ten plants were used for each of the four repetitions for a total of 320 plants.

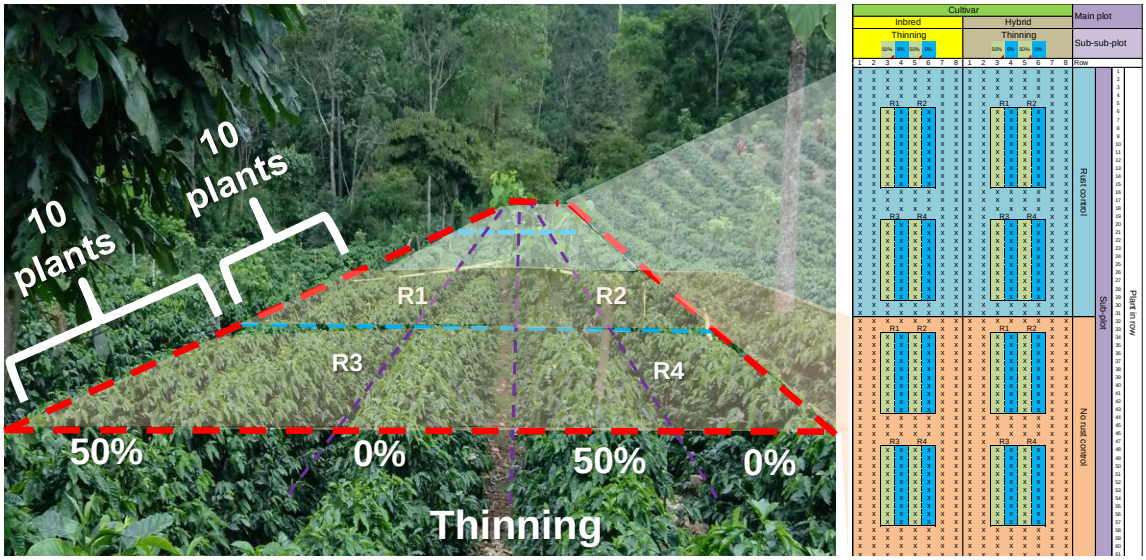


Figure 8. Experimental design layout on the farm.
Left photograph shows an aerial view of the inbred plots with the repetitions for each rust control treatment. The right scheme represents the overall layout for cultivars (main plots), rust control (sub-plot), and fruit thinning (sub-sub-plot) treatments. Reprinted with permission from Fabián Echeverría-Beirute, Seth C. Murray, Patricia Klein, *et al*. 2017. Rust and Thinning Management Effect on Cup Quality and Plant Performance for Two Cultivars of *Coffea arabica* L. Journal of Agricultural and Food Chemistry. Copyright © 2017 American Chemical Society.

2.2.2 Plant overall performance evaluation

Through the year, the coffee plants had different developmental stages. Flowering happened between March to May. The fruit ripening occurred from October to December. Before June, almost no significant rust was observed because of good agronomical conditions. Therefore, each plant was evaluated every two months from June 2015 through January 2016; as expected to be the worst months of the disease with relationship to yield. The evaluation was performed in six branches per tree. Each branch was selected and labelled according to the trunk height (superior [top] and medium [half]) and cardinal direction (West [W], North [N] and East [E]). The parameters

evaluated per branch were total number of leaves (TL), number of leaves with coffee leaf rust (CLR), percentage of lesions (NL) per leaf, and percentage of those lesions sporulating (NE). The rust incidence (RI) was obtained by dividing the average of all branches CLR per TL [RI = CLR/TL]. Rust severity (RS) was obtained by multiplying the average RI of all branches by NL and dividing by 100 [RS = (RI x NL)/100]. The rust sporulation (RE) was obtained by dividing NE by NL (RE = NE/NL). All RI, RS, and RE were interpreted as percentages (%). A subjective overall condition (OC) was also generated (by F.E.-B.) considering a visual aspect of the general plant physiology and health on a scale of 1 to 5, where "1" represented bad and "5" excellent condition. The total harvest (TH) per plant (fresh weight) was obtained by weighing all manually harvested fruits of each plant in every optimal ripening. Five harvests were made every fifteen days, beginning in October and ending in December (200 to 240 days after flowering). The TH data (in kg) was summed into an accumulated yearly production per plant for further statistical analysis. Data per month obtained for TL, OC, RI, RS, and RE was used for the statistical analysis of the stress and management interactions.

2.2.3 Fruit sample preparation

At the peak of the harvest season (when ≥ 35% of the accumulated harvest was in the optimal stage), the berries were harvested from all the plants belonging to each repetition and management treatment, weighed, later bulked, and mechanically depulped by a wet milling process. The mucilage was washed 24 hours later and the beans were dried up to 11% moisture as determined by a portable moisture tester (G600i, AGRI,

GEHAKA). Later, the green beans were mechanically recovered from the parchment. Thirty-two dried green bean samples of 300 grams/sample, corresponding to the two cultivars by two thinning by two rust control treatments, with four biological repetitions, were vacuum-sealed in plastic bags and transported to Texas A&M University (USA).

2.2.4 Quality profile evaluation

Each sample was roasted and cupped by Songer & Associates, Inc. (Colorado, USA) following the SCAA standard protocols [88]. Roasting final degree varied from 55-60 ColorTrack (45 Agtron). The data reported included the clean cup, acidity, body, overall, and final scoring of each sample. The same samples were used to analyze the aroma and flavor profile using the WCR Sensory Lexicon [159] at the Sensory Laboratory in the Animal Science Department at Texas A&M University (TAMU) by four trained judges (authorization statement IRB 2015-0423M). The evaluation of the 32 samples was carried out over five sessions; between six to seven random samples were analyzed. The order of presentation was randomized among judges and sessions. The attributes profile of each sample was defined by the consensus of the judges' intensity perception to each of the 60 attributes found.

2.2.5 Volatile compound analysis

The profiles of volatile compounds were obtained with the method described by Bertrand, et al. [111], and performed in the Flavor Chemistry Laboratory at the Department of Animal Science at TAMU. Briefly, 2 grams of roasted beans were ground and placed

in a 2 ml hermetically sealed glass flask, which corresponded to a headspace of ⅓ of the sampling flask. The flasks were placed in a thermostatically regulated oven at 50°C for 15 min until the sample headspace equilibrium was reached. Alkane standards (C7 to C30; Catalog #49451-U; Sigma Aldrich, St. Louis, MO, 63103) were run prior and after the injection of the experimental samples to verify the retention times of sample alkanes, in order to validate the compounds obtained in the mass spectra (MS). The volatile compounds were collected by Solid Phase Micro extraction (SPME) using a 75 µm Carboxen/polydimethylsiloxane (CAR-PDMS) fiber (Supelco 504831, Sigma-Aldrich, St. Louis, MO) in the injector port, followed by Gas Chromatography-Mass Spectrometry (GC-MS) (Agilent Technologies 7920 series GC and Agilent Technologies 5975 series MS, Santa Clara, CA) analysis over 45 minutes. The helium carrier flow was at 1.5 ml/min and used an AromaTrax® system (MicroAnalytics, Inc., Round Rock, TX, USA), with dual sniff ports for characterization of aromatics. The volatile compounds were identified by comparing their MS to those of the Wiley Chemical Library database and by their retention times with those of standard compounds [160]. The total content of the volatiles of each headspace analysis was defined by integrating the peak areas of the 154 key odorants identified. The measured units are relative abundance with units of total ion count.

2.2.6 Statistical analysis

The normality of the data was tested through goodness of fit. The data for all parameters was non-normal, since they came from different sources of variability. The

transformation of the data did not fit Gaussian or even Johnson's normality measures. A generalized linear model which is less susceptible to normal distribution deviations was used for analysis [161]. The generalized linear model was constructed to analyze the plant overall performance data using a fixed factorial model of each cultivar (G), fruit thinning (T), and rust control (C), used per each parameter according to:

$P=\mu_G + \mu_T + \mu_G * \mu_T + \mu_C + \mu_G*\mu_C + \mu_C*\mu_T + \mu_G*\mu_T*\mu_C + \mu_e$, where P represents the phenotype for each trait, each μ the means of the corresponding variables, and μ_e the unexplained error. The statistical model was used to describe the variance of the components by repeated measurements (each month) using multiple way analysis of variance (MANOVA). The MANOVA considered the interaction of $P(\mu_{th}*\mu_{cd}*_{month})$, where μ_{th} represents the trunk height, μ_{cd} the cardinal direction, and each plant served as a replication (R). The cumulative average was adjusted by the least square means (LSM). The LSM were used to identify significant differences between the interactions of genotype, thinning, and rust control, using Tukey HSD test with a significant difference at $p \leq 0.05$. Simple *t*-test analysis at $p \leq 0.05$ was used to validate the ANOVA analysis of the sensory and chemical data according to cultivar (G), fruit thinning (T), or rust control (C) without interactions. Correlation analysis between plant overall performance and quality data was performed using Pearson for normal data (TL and OC) and Spearman's correlations for the non-normalized data. Probability was set to $p \leq 0.01$, and the associations of plant overall performance with sensory or chemical data, were considered to plot the LSM in a principal component analysis (PCA), in order to describe relationships among variables and differences among the management

treatments. All statistical analyses were performed using JMP Pro 12.0.1. (SAS Institute Inc. USA) software package.

2.3 Results

2.3.1 Overall plant performance profiles

The Figure 9 shows the effect of the variables used to measure the traits leading to overall plant performance. The variance of each trait was affected in the first order by the month (60%), followed by rust control (20%), the cultivar (8%), the branch height of the plant where the data was collected (7%), the fruit thinning treatment (3%), which particular plant the data was collected from (2%), and finally the cardinal orientation of the branch sampled (1%). Although all these main effects existed, the coefficient of determination (R^2) showed that for each trait, between 66% and 75% of the variation could be explained by these effects; therefore, the least square means (LSM) were used to adjust the means across all other covariates to better estimate the statistical significance of the treatments in the model.

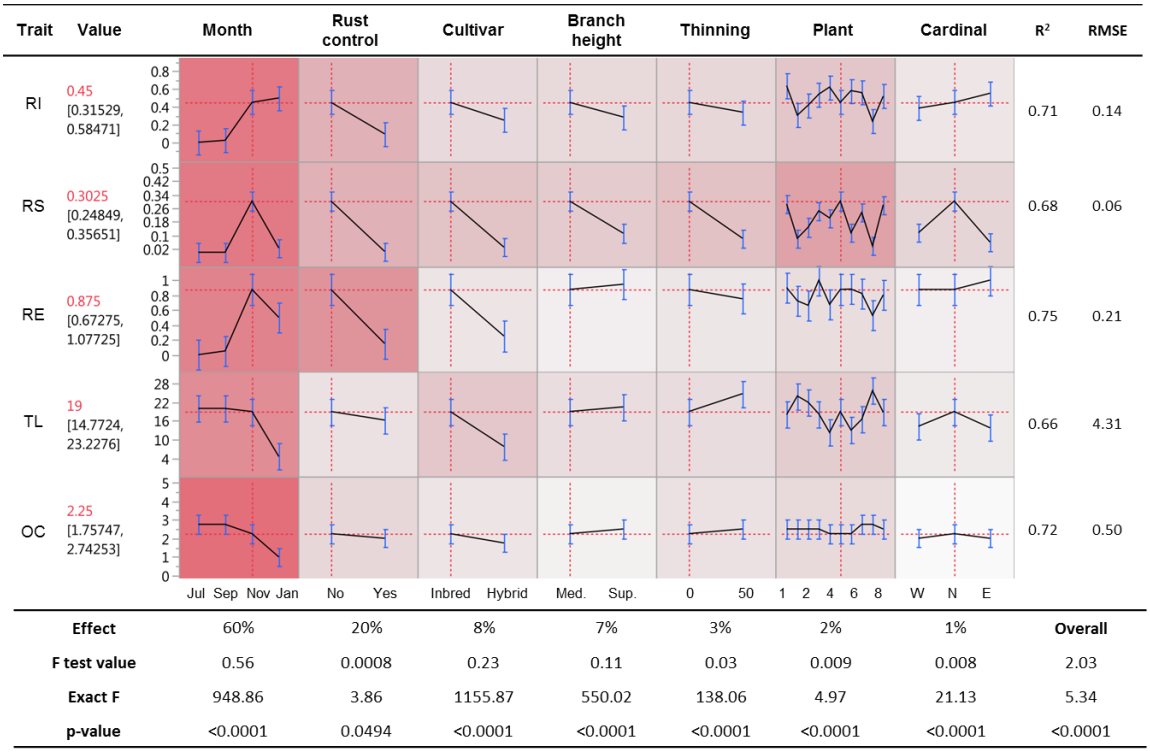
The analysis of rust incidence (RI), rust severity (RS), rust sporulation (RE), and total leaves (TL) on a monthly basis, revealed different patterns between cultivars and management treatments (rust control and thinning) (Figure 10). For both cultivars, the most observable difference was between rust control treatments. Under the rust controlled treatments, RI slowly increased until November, reaching a level of 5-10%. When the plants did not receive a rust control treatment, RI showed an increase from almost 0% in July, to a maximum level of 20 to 45% between November and January.

The rust severity (RS) had a similar increasing pattern as RI for both cultivars when they did not receive a rust control treatment. The RS reached maximum levels of 10 to 15% between November to January. Rust control had the largest effect on RE. When plants were not treated for rust, a dramatic increase in RE was observed from September through January, reaching levels between 50 to 80% in both cultivars. The plants with rust control, showed a slight increase (5 to 10%) in RE between July to September, but declined to nearly 0% by the final evaluation in January. The TL was relatively uniform in both cultivars throughout the season, with a slight decrease by January.

The LSM of RI, RS, RE, TL, OC, and TH, were used to examine the effect of each treatment (rust control and fruit thinning) on the overall response of each cultivar (Table 3). Under rust control and fruit thinning treatments, RI, RS, and RE, were significantly lower in both cultivars. For both cultivars, RI was ~12% lower under rust control and ~4% lower under thinned treatments. Overall, RI was lowest in hybrid plants treated for rust and thinned to 50% fruit yield (3%), and highest in the inbred when plants were not treated for rust or thinned to reduce fruit yield (21%). Additionally, for both cultivars, RS was less than 1% in rust control treated plants, whereas under thinning treatments, RS was 2% lower than under non-thinned treatments. RS was ~5% for both cultivars under no rust control and no thinning. Rust control prevented ~27% of RE overall, and thinning treatments reduced RE by ~2% when compared to non-thinned treatments. The RE was ~9% lower in the hybrid with no rust control when compared to the inbred under the same treatment.

TL varied significantly between cultivars, with a general average of four more leaves per branch in the inbred compared to the hybrid. The inbred reached a maximum of ~18 TL under rust control and thinning treatments, but had ~15 under rust control and no thinning treatments (Table 3). The hybrid obtained a maximum of ~13 TL almost independently of the treatments, with exception of the no rust control and non-thinned treatments, which had an average of ~12 TL. When comparing overall rust control and thinning treatments, the plants that had rust control showed less than one more leaf per branch than no rust control, and one more leaf on average than those that had been thinned. Leaf drop at the end of the harvest season was expected due to CLR, wind, and hand picking, therefore, TL attached to the plant was considered as an indirect expression of health condition.

The subjective evaluation of OC differed due to rust control, thinning, and cultivar (Table 3), especially during the two last months of evaluation (Figure 10). The rating for OC was 0.2 points lower in plants without rust control and for those plants that were not thinned. The difference between cultivars was 0.1 points. However, statistically, the major difference observed was between rust controlled and thinned plants from the inbred, and the plants without rust control or thinning in the hybrid, which reached the maximum level of 2.7 points and the minimum level of 1.9 points, respectively.



	Month	Rust control	Cultivar	Branch height	Thinning	Plant	Cardinal	Overall
Effect	60%	20%	8%	7%	3%	2%	1%	
F test value	0.56	0.0008	0.23	0.11	0.03	0.009	0.008	2.03
Exact F	948.86	3.86	1155.87	550.02	138.06	4.97	21.13	5.34
p-value	<0.0001	0.0494	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Figure 9. The effect of the sources of variance for plant performance traits.
Values in the y-axis represent means (red) and range (black), and the variables are shown in the x-axis. Definitions for each variable can be found in Materials and Methods. Bars in each graph represent standard error (SE). Coefficient of determination (R^2) and root mean square error (RMSE) for each trait are presented. The overall effect of each variable is displayed below the graphs, along with the corresponding statistical analysis according to the MANOVA. Reprinted with permission from Fabián Echeverría-Beirute, Seth C. Murray, Patricia Klein, *et al*. 2017. Rust and Thinning Management Effect on Cup Quality and Plant Performance for Two Cultivars of *Coffea arabica* L. Journal of Agricultural and Food Chemistry. Copyright © 2017 American Chemical Society.

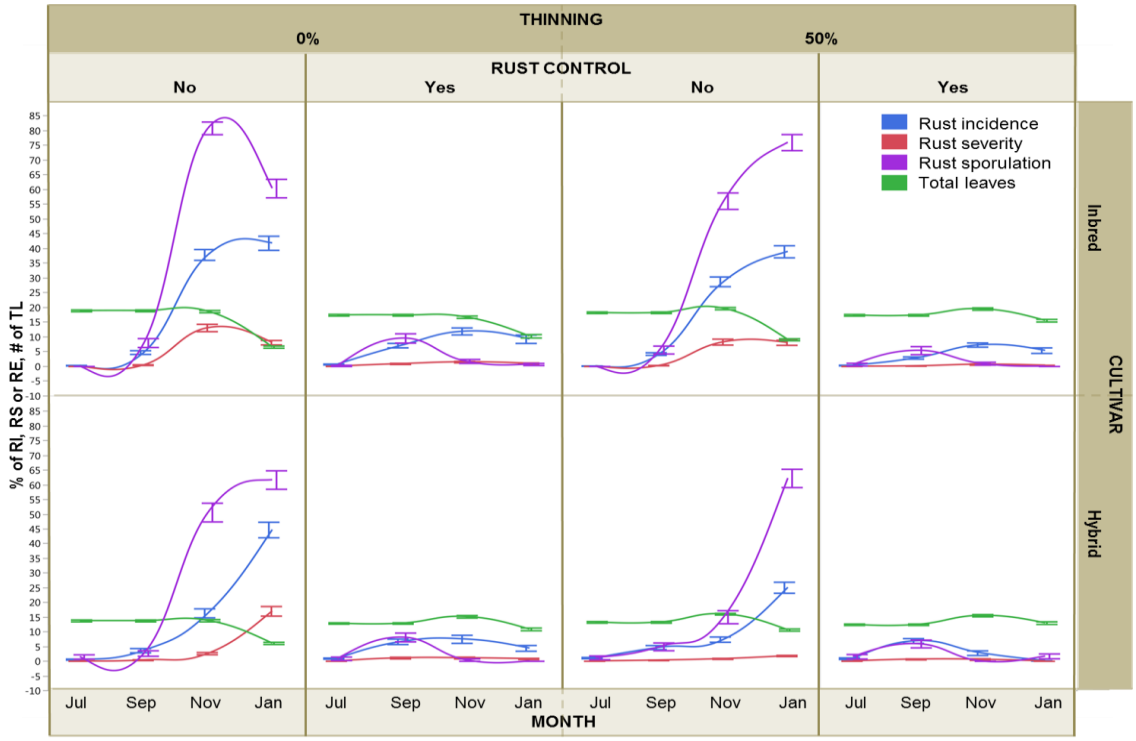


Figure 10. Monthly mean effect of rust control (no or yes) and fruit thinning (0% or 50%) on rust incidence (RI), rust severity (RS), rust sporulation (RE), and total leaves (TL) in both the inbred and hybrid cultivars.
Error bars represents standard error (SE). Reprinted with permission from Fabián Echeverría-Beirute, Seth C. Murray, Patricia Klein, *et al*. 2017. Rust and Thinning Management Effect on Cup Quality and Plant Performance for Two Cultivars of *Coffea arabica* L. Journal of Agricultural and Food Chemistry. Copyright © 2017 American Chemical Society.

As expected, the TH showed significantly lower yield in plants subjected to fruit thinning, with 1.1 kg (64% of total) and 1.4 kg (62% of total) differences for the inbred and hybrid, respectively (Table 3). Rust control treated plants did not have a significant increase in TH, but an average increase of 0.1 to 0.2 kg fruits per plant in the inbred and hybrid, respectively.

Table 3. Overall mean effect of rust control (no or yes) and fruit thinning (0% or 50%) on plant performance traits (RI, RS, RE, TL, OC and TH) in both the inbred and hybrid cultivars.

Treatment	Cultivar	Thinning	Rust control	RI	RS	RE	TL	OC	TH
1	Inbred	0%	Yes	7.0% c	0.8% a	3.0% a	15.4 c	2.3 cd	2.9 abc
2	Inbred	0%	No	21.0% f	5.4% c	34.0% e	15.8 bc	2.3 d	2.9 abc
3	Inbred	50%	Yes	4.0% ab	0.3% a	2.0% a	17.5 a	2.7 a	1.9 c
4	Inbred	50%	No	18.0% e	4.2% b	34.0% d	16.2 b	2.3 cd	1.8 c
5	Hybrid	0%	Yes	5.0% b	0.7% a	2.0% a	12.8 d	2.3 c	3.8 a
6	Hybrid	0%	No	16.0% e	5.0% c	29.0% c	11.7 e	1.9 e	3.7 ab
7	Hybrid	50%	Yes	3.0% a	0.3% a	2.0% a	13.3 d	2.4 b	2.5 bc
8	Hybrid	50%	No	10.0% d	0.7% a	21.0% b	13.2 d	2.3 cd	2.2 abc
Overall effect		Rust control (yes)		-11.5%	-3.3%	-27.3%	0.5	0.2	0.1
		Fruit thinning (50%)		-3.5%	-1.6%	-2.3%	1.1	0.2	-1.2
		Cultivar (hybrid)		-4.0%	-1.0%	-4.8%	-3.5	-0.1	0.6

Letters next to the value represents least significant difference (LSD) at $p \leq 0.05$. Overall effect at the bottom represents the difference between management practices or cultivar change. Reprinted with permission from Fabián Echeverría-Beirute, Seth C. Murray, Patricia Klein, *et al*. 2017. Rust and Thinning Management Effect on Cup Quality and Plant Performance for Two Cultivars of *Coffea arabica* L. Journal of Agricultural and Food Chemistry. Copyright © 2017 American Chemical Society.

2.3.2 Sensory and volatile compound profiles

The SCAA sensory lexicon showed differences in the scores for clean cup, acidity, body, and overall attributes, according to the treatments, but not in the final score, which had an average of 85.54 ± 2.29 (data not shown). The clean cup and overall scores were 0.2 and 0.3 points lower, respectively, when the plants had rust control (Figure 11, Appendix 1). The acidity score was 0.3 points higher in beans harvested from thinned plants. The score for body was 0.1 points higher in the inbred. The body attribute was the only trait that also displayed significant G x M interaction in the SCAA sensory lexicon ($p = 0.0122$, $R^2 = 0.42$; data not shown).

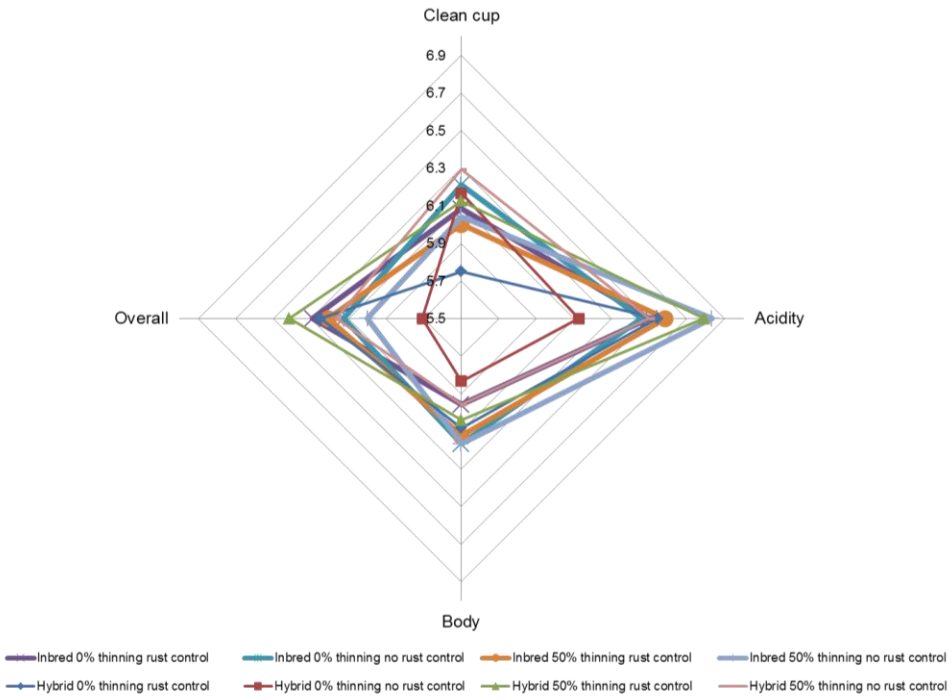


Figure 11. Effect of rust control (no or yes) and fruit thinning (0% or 50%) on the interactions of statistically significant sensory attributes using the SCAA lexicon in both the inbred and hybrid cultivars.
Significant attributes ($p \leq 0.05$) were identified using the SCAA Sensory Lexicon (Appendix 1). Sensory attributes scored between on a scale from 5 to 7. Reprinted with permission from Fabián Echeverría-Beirute, Seth C. Murray, Patricia Klein, *et al*. 2017. Rust and Thinning Management Effect on Cup Quality and Plant Performance for Two Cultivars of *Coffea arabica* L. Journal of Agricultural and Food Chemistry. Copyright © 2017 American Chemical Society.

The beverage quality analyzed by the WCR Sensory Lexicon revealed small changes in intensity for 6 out of 60 different attributes (Figure 12, Appendix 2). Under thinning treatments, the cardboard aroma (S19) and burnt flavor (S33) significantly decreased by 0.6 points. When plants were not treated for rust, musty dusty aroma (S23) decreased 0.8 points. The overall impact flavor (S26) was one point higher in the inbred. The mouth drying astringent flavor (S29) was 0.6 points lower in the hybrid. Stale aroma (S24) statistically decreased with thinning and rust control in the hybrid, but increased in

beans from thinned and rust controlled treatments in the inbred. Stale aroma (S24) was the only attribute with significant G x M interaction in the WCR sensory lexicon (p = 0.0033, R^2 = 0.55; data not shown).

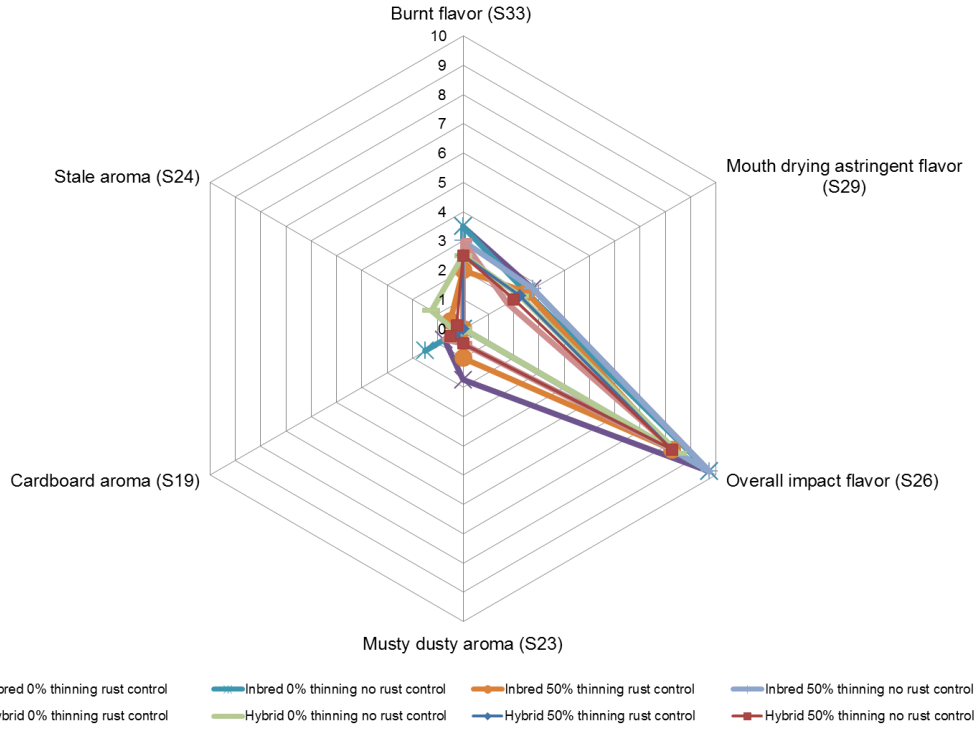


Figure 12. Effect of rust control (no or yes) and fruit thinning (0% or 50%) on the interactions of statistically significant sensory attributes using the WCR lexicon in both the inbred and hybrid cultivars.
Significant attributes ($p \leq 0.05$) were identified using the WCR Sensory Lexicon (Appendix 2). Sensory attributes scored between on an intensity scale from 0 to 10. Reprinted with permission from Fabián Echeverría-Beirute, Seth C. Murray, Patricia Klein, *et al*. 2017. Rust and Thinning Management Effect on Cup Quality and Plant Performance for Two Cultivars of *Coffea arabica* L. Journal of Agricultural and Food Chemistry. Copyright © 2017 American Chemical Society.

Among the chemical volatile compounds, 18 out of 154 were statistically significant (Figure 13, Appendix 3). The volatile compounds 2-propionylfuran (V4), 1,2,3,4-tetrahydronaphthalene-5-carboxaldehyde (V5), ketole (V6), 4'-hydroxy-3'-

methylacetophenone (V21), pentanal (V40), guaiacol (V41), and 1-furfurylpyrrole (V73) increased with thinning. Coumarone (V58), 2-furoylacetonitrile (V77), and methanethiol (V89) were higher in beans from plants without rust control. Furfural acetone (V19) was lower in beans from the hybrid and increased with rust control, while phenyl acetate (V55) was higher in beans from the hybrid and when no rust control was made. Vinylveratrole (V25), difurfuryl ether (V37), pyrrole (V105), 2-formyl-1-methylpyrrole (V114), and camphor (V119), were higher in the inbred than in the hybrid. Acetaminophen (V82) was significantly reduced in beans from the hybrid under thinning and rust control treatments. To our knowledge, acetaminophen has not previously been reported in coffee, and remarkably, was associated with an increase in RI, RS, and RE, and with a decrease of OC (Figure 14). Phenyl acetate (V55), coumarone (V58), and pyrrole (V105) showed significant G x M interaction ($p = 0.0042$, $R^2 = 0.61$; $p = 0.0487$, $R^2 = 0.37$; and $p = 0.0224$, $R^2 = 0.37$, respectively; data not shown).

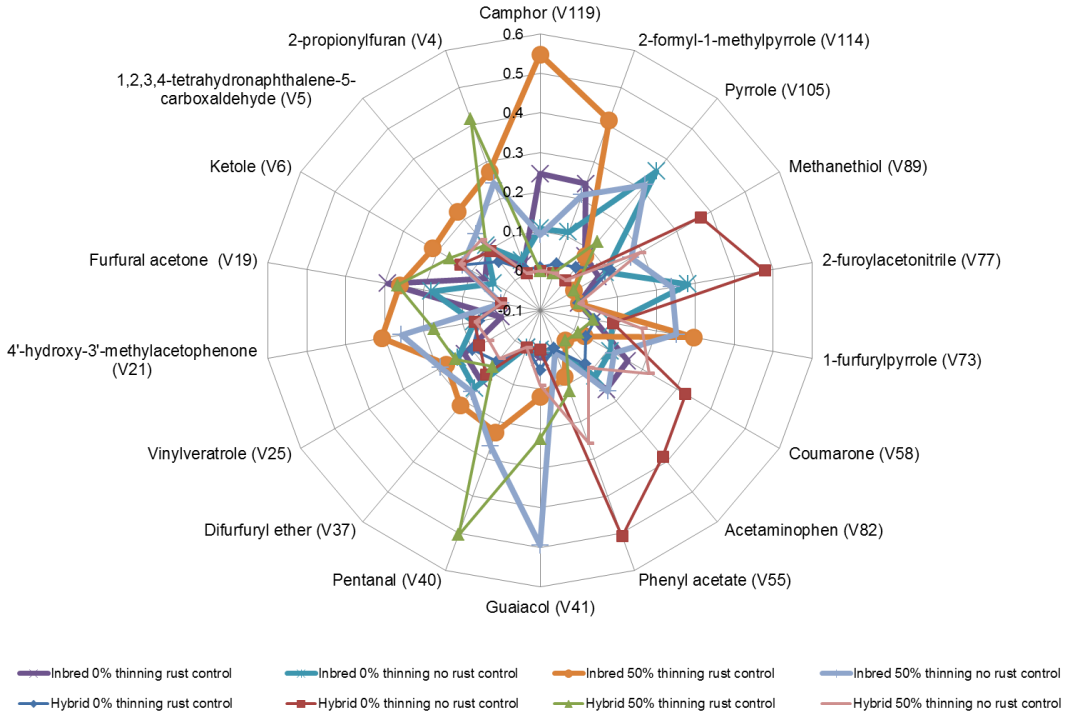


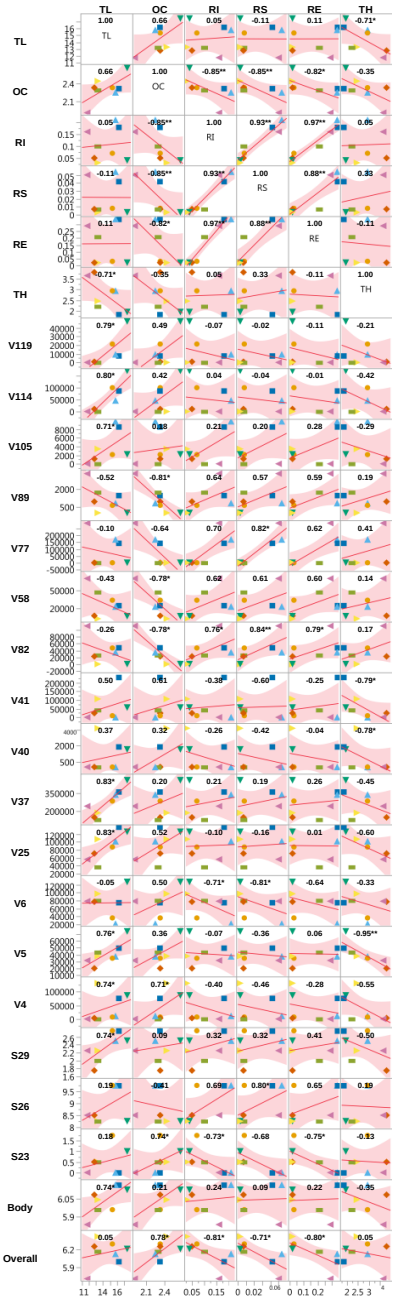
Figure 13. Effect of rust control (no or yes) and fruit thinning (0% or 50%) on the interactions of statistically significant ($p \leq 0.05$) volatile compounds in both the inbred and hybrid cultivars (Appendix 3).
Each volatile compound is presented as a percentage of abundance between treatments. Reprinted with permission from Fabián Echeverría-Beirute, Seth C. Murray, Patricia Klein, *et al*. 2017. Rust and Thinning Management Effect on Cup Quality and Plant Performance for Two Cultivars of *Coffea arabica* L. Journal of Agricultural and Food Chemistry. Copyright © 2017 American Chemical Society.

The correlations of the different variables revealed significant associations ($p \leq 0.05$) between overall plant performance and bean quality (Figure 14). The body attribute was significantly correlated with TL (r = 0.74) using the SCAA protocol, while the overall attribute was correlated (r = 0.78) with an increase of OC, but negatively correlated to RI (r = -0.81), RS (r = -0.71) and RE (r = -0.80). With the WCR Sensory Lexicon, the attribute "musty dusty" significantly increased with OC (r = 0.74), but decreased linearly with RI (r = -0.73) and RE (r = -0.75). Furthermore, the overall

impact flavor, increased with RS (r = 0.80), while the mouth drying astringent flavor increased with TL (r = 0.74).

From the volatile compounds, significant correlations ($p \leq 0.05$) were found (Figure 14). The volatiles 2-propionylfuran (V4), 1,2,3,4-tetrahydronaphthalene-5-carboxaldehyde (V5), vinylveratrole (V25), difurfuryl ether (V37), pyrrole (V105), 2-formyl-1-methylpyrrole (V114), and camphor (V119), were positively correlated with TL. Methanethiol (V89), coumarone (V58), and acetaminophen (V82), decreased with OC. The volatile 1,2,3,4-tetrahydronaphthalene-5-carboxaldehyde (V5), pentanal (V40), and guaiacol (V41), were negatively correlated with TH. Ketole (V6) was negatively correlated with RI and RS, while 2-furoylacetonitrile (V77), increased linearly with RS. Only 1,2,3,4-tetrahydronaphthalene-5-carboxaldehyde and acetaminophen were highly significantly correlated ($p \leq 0.01$) to TH and RS, respectively.

All highly significant ($p \leq 0.01$) plant performance and quality correlated variables, were included in a principal component analysis (PCA) to identify potential associations (Figure 15). This resulted in 87% of the total variation being described by the first two principal components.



Variables' nomenclature

Plant overall performance:

TL= total leaves / branch
RI= rust incidence
RS= rust severity
RE= rust sporulation
TH= total harvest / plant
OC= overall condition

Sensory:

WCR lexicon	**SCAA lexicon**
S23 = musty dusty aroma	Body
S26 = overall impact flavor	Overall
S29 = mouth drying astringent flavor	

Volatiles:

V4 = 2-propionyfuran
V5 = 1,2,3,4-tetrahydro-naphthalene-5-carboxaldehyde
V6 = ketole
V25 = vinylveratrole
V37 = difurfuryl ether
V40 = pentanal
V41 = guaiacol
V58 = coumarone
V77 = 2-furoylacetonitrile
V82 = acetaminophen
V89 = methanethiol
V105= pyrrole
V114= 2-formyl-1-methylpyrrole
V119= camphor

Figure 14. Scatterplot matrix representing the correlation of plant performance, sensory attributes, and volatile compounds.
Significant correlations ($p \leq 0.05$) are indicated with asterisk (*). Highly significant correlations ($p \leq 0.01$) are indicated with double asterisk (**). Reprinted with permission from Fabián Echeverría-Beirute, Seth C. Murray, Patricia Klein, *et al*. 2017. Rust and Thinning Management Effect on Cup Quality and Plant Performance for Two Cultivars of *Coffea arabica* L. Journal of Agricultural and Food Chemistry. Copyright © 2017 American Chemical Society.

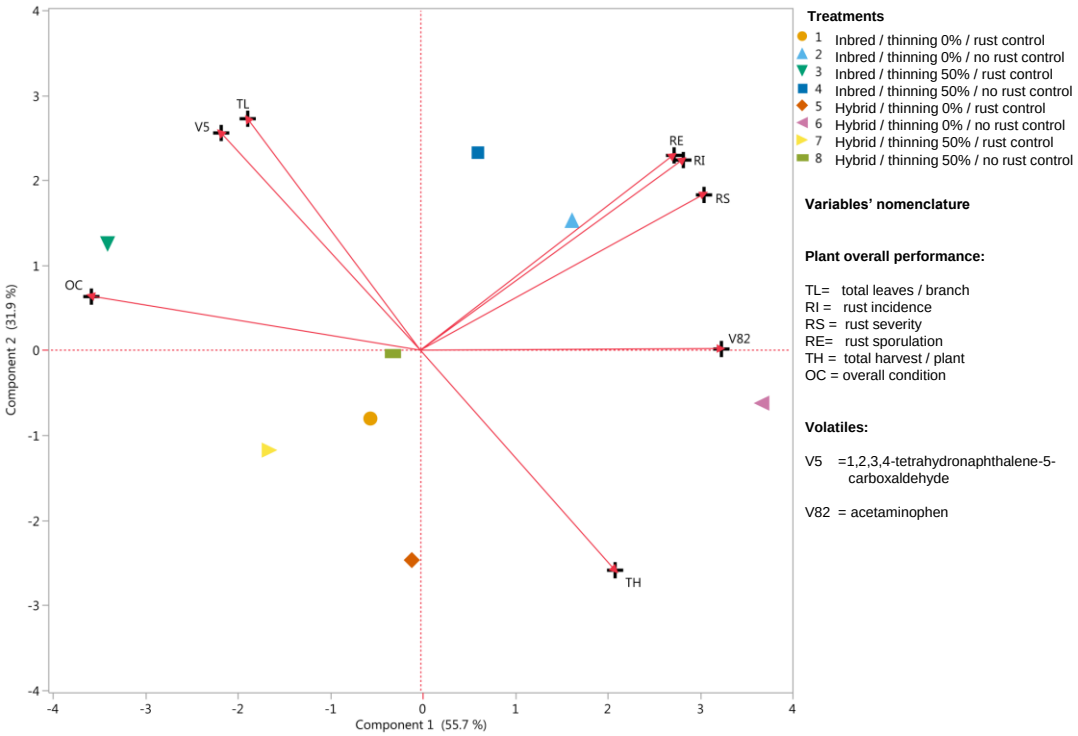


Figure 15. Principal component analysis (PCA) biplot for the treatments, according to the interaction of plant performance and chemical volatile analysis. Reprinted with permission from Fabián Echeverría-Beirute, Seth C. Murray, Patricia Klein, *et al*. 2017. Rust and Thinning Management Effect on Cup Quality and Plant Performance for Two Cultivars of *Coffea arabica* L. Journal of Agricultural and Food Chemistry. Copyright © 2017 American Chemical Society.

2.4 Discussion

The effect of the interaction of rust control and fruit thinning management practices were evaluated on six aspects of plant performance, bean quality attributes, and volatile compounds in two cultivars of coffee. The thinning treatment effectively reduced the total harvest ~37% when compared with the non-thinned treatment in both cultivars, but was less than the expected 50% reduction. Since the yield was evaluated

by fresh fruit weight, possible reasons for this difference of ~13% could be that the number of flowers manually removed wasn't precisely 50% under the thinning treatments, or that either the fruits or beans compensated by increasing in weight compared to non-thinned plants. Studies by Vaast, et al. [75] and Bote and Jan [153] did not find significant differences in bean size after thinning coffee plants, suggesting that other effects such as dry weight and density may explain variances in the bean yield. Bote and Jan [153] reported an increase of ~21-22% (2 mg/bean) in dry bean weight, and 11-14% (0.5 cm) in bean diameter, when a 75% fruit thinning treatment was performed, suggesting that significant changes in bean morphology may be achieved under higher thinning treatments (more than 50%). In our study, dry weight per bean did not significantly change (data not shown); therefore, the small differences in the expected yield could be caused by other factors, such as pulp weight change, which we did not measure.

Rust control (cyproconazole and epoxiconazole) had an effect of controlling 12% of the rust incidence (RI), 3% of severity (RS), and 27% of the sporulation (RE) on both cultivars, whereas without rust control, RI, RS, and RE, reached 21%, 5%, and 34%, respectively. The active ingredient in these fungicides is a steroid demethylation inhibitor, which blocks the C14-demethylase enzyme in the fungus. This enzyme plays a role in the production of sterols such as ergosterol, required for membrane structure and function, and therefore, for normal morphology, cell growth, and reproduction of the fungus [162]. Even though the sprays were applied according to the label directions, some residual lesions and spores were able to recover and reinitiate a subsequent infection

cycle, similar to what takes place in a commercial setting. Therefore, the best recommendation is to continue to implement integrated pest management for the control of CLR and to include resistant or tolerant cultivars where possible (e.g., derived from sarchimors, catimors, and introgressions from *C. liberica* or *C. canephora*).

The rust control treatments efficiently controlled CLR and protected the harvest from non-significant estimated losses of around 0.1 kg per tree (4.2%) in the inbred and 0.2 kg (6.5%) in the hybrid, equivalent to around 11 to 22 kg of green bean coffee per hectare, respectively. Under commercial conditions, factors including other management practices, plant nutrition, and weather, are sources of variability that will generally result in larger yield loss. For example, in 2012, Costa Rica had an extreme CLR event with more than 60% of the coffee growing areas damaged, 16% of the plants had severe leaf fallout, and 13% of the harvest during that period was lost [26]. Likely causes of this event were 20% more time at temperatures that favor CLR infections (20-25°C), in combination with lower than average fertilization applications on farms due to high costs [26]. Coffee plants stressed by other conditions are more susceptible to CLR, which is why both nutrition and rust control are emphasized in dealing with this disease. In our experiment, plants in all treatments received equal nutrition; therefore, the only management practices from a commercial production standpoint that we modified, was rust control.

Given the value of coffee at ~$0.61/kg and the cost of production at ~$0.53/kg in Costa Rica [Marco Araya, Instituto del Café de Costa Rica (ICAFE)], personal communication), every hectare (Ha) at a medium production level, has an expected

return of ~$560 per year. Therefore, it is not economically feasible to perform three or more extra treatments for rust control at a cost of ~$55 per application/Ha (not including the cost of human labor). The case is even worse for low yielding farms, which have costs of production of ~$0.64/kg [163]. Therefore, with similar rust pressure, there would need to be an economic incentive for the protection that rust control provides, to not only preserve bean quality, but also bean yield. Decisions on the economic feasibility of rust control treatment should be primarily focused on how much yield loss will be prevented and how overall bean quality throughout the harvest season will be affected. This study used the highest quality from the overall harvest to examine the effect of the treatments on bean quality, since it is known that quality (sensory and bean composition), is associated with ripening and cultivar, and that CLR causes premature ripening of berries that produce poor-quality coffee beans.

The hybrid showed more tolerance to CLR than the inbred, since less RI (4% difference), RS (1%), and RE (5%) was observed, as well as a delayed appearance of severe symptoms and leaf drop-off (Figure 10). Although the hybrid showed some level of tolerance, the fungus was able to establish, spread spores, and increase disease during the last month of evaluation. The tolerance of the hybrid likely was derived from the wild accession (Ethiopia 531) in its background. As a co-evolutionary strategy, tolerance is more likely to be fixed in the populations in the areas of origin, as opposed to breeding selection programs outside the center of origin that inadvertently fix qualitative traits, such as likely is the case of the inbred (Catuai vermelho IAC 144) [164]. The overall productivity of the hybrid was 20% higher than the inbred, and can be explained by the

heterosis of the hybrid, which had higher levels of vigor and productivity, but could also be partially due to the fact that the hybrid was one year older than the inbred.

The sensory analysis of the coffee samples showed variability in some sensory attributes used to evaluate beverage quality with both the SCAA and WCR Sensory Lexicons. Interestingly, even though both cultivars shared several attributes and each obtained a final SCAA score that would rate them as an excellent specialty coffee, they had different sensory and volatile profiles (i.e. body attribute, overall impact flavor, mouth drying astringent flavor, furfural acetone, phenyl acetate, vinylveratrole, difurfuryl ether, pyrrole, 2-formyl-1-methylpyrrole, and camphor) demonstrating that independent components of quality were genotype (G) and management (M) dependent under the same environment (E). Quality can clearly be affected from the interaction of G, M, and/or E. de Oliveira Fassio, et al. [165] showed from ten descriptors obtained from a SCAA quality evaluation, that only flavor and acidity were necessary to differentiate between nine cultivars (G) and two environments (E). The interaction of G x E was responsible for the final score, but they also identified some cultivars that exhibited lower variance between environments (less G x E interaction). Furthermore, their sensory analysis showed no correlations between sensory and bioactive compounds such as caffeine, trigonelline, and 5-caffeoylquinic acid (5-CQA), across cultivars, revealing that those compounds aren't consistent markers of quality or as impacted by G x E interactions; as supported by other studies as well.

The correlation analysis (Figure 15) revealed that only 1,2,3,4-tetrahydronaphthalene-5-carboxaldehyde and acetaminophen were highly significant to

yield and rust severity, respectively. Acetaminophen and paracetamol are common names of N-(4-hydroxyphenyl)-acetamide (V82), commercially produced since 1953 because of its antipyretic analgesic effect on animals [166]. Phenylamine biosynthesis is involved in the shikimate pathway related to the production of other aromatic amino acids like tyrosine and tryptophan; and the phenylpropanoid pathway, which regulates the synthesis of caffeic acid, salicylic acid, and lignin [167-168]. Moreover, phenylalanine is required for the biosynthetic pathway of anthocyanin , related to other pathways linked to flavonoids and plant hormones [169] demonstrated that stress induces phenylpropanoid metabolism in order to enhance abiotic and biotic defense responses in many plants. Since the phenylalanine molecule is demonstrated to be a precursor of at least 14 volatile compounds in coffee, we hypothesize that the production of acetaminophen (V82) was also produced as a transformation of phenylalanine (and/or derivative compounds) by the roasting process.

Acetaminophen was 5X and 6X higher under no thinning and no rust control treatments in the hybrid, respectively, and 2X higher in the inbred under these stress conditions (Figure 14). It was also linearly correlated with RI, RS, and RE (Figure 15); indicating that higher abundance can be related to the defense response of the coffee plant against CLR. To our knowledge, this is the first report that acetaminophen can be produced in roasted coffee beans. This finding might also support the necessity of studies measuring G x M stress response and quality, and the use of more sensitive detection methodologies such as GC-MS and SPME. That we are the first to identify this

compound in coffee exemplifies the enormous complexity of the physiology, genetics, and chemistry involved in coffee quality.

We conclude that the management of fruit yield and coffee leaf rust disease (CLR) affected coffee cup quality and volatile composition under our environmental conditions. Future studies under different conditions should focus on explaining and validating the intricate network of effects between G, M, and E on cup quality and plant performance.

CHAPTER III

BEAN TRANSCRIPTOME AND VOLATILES RELATED TO FATTY ACIDS AFFECTED UNDER RUST (*HEMILEIA VASTATRIX* BERK. & BR) AND YIELD STRESSES IN COFFEE (*COFFEA ARABICA* L.)

Stress is one of the major problems induced by coffee leaf rust (CLR), which is caused by *Hemileia vastatrix* Berk. et Br. The severity of CLR damage increases during the higher bearing years. Transcriptome literature has previously focused on the physiological response of leaves to CLR but not on fruits or the quality of the beverage precursors. This study evaluated the effect of CLR control and fruit thinning treatments on the gene expression of immature and mature coffee beans in two CLR susceptible cultivars of *Coffea arabica*. The RNA-seq analysis was oriented to the functional description of ripening time and the treatment effects on green bean volatile precursors identified by SPME-GC/MS. A total of 471 differentially expressed genes (DEGs) were grouped into 19 gene ontology (GO) functional categories related to the treatments, maturity stages, and cultivars. The enriched metabolic pathways related to the DEGs revealed differences between the management practices and the physiology of the plant by genotype. A higher number of DEGs were found in the immature stage where synthesis of fatty acids and carbohydrates were most active. Structural modifications and accumulation of metabolites in the cell wall differed between treatments, revealing activation of metabolism caused by stress. Stress changed both gene expression and volatile profiles in pathways especially related to unsaturated fatty acid (linoleic and

oleic acids) metabolism; reflecting an increase in their oxidized volatile forms nonanal, (E,E)-2,4-Decadienal, and (E)-2-Decenal. The overall interaction of rust control and fruit thinning management showed that stress influences the bean's defense response and the chemical composition in a cultivar dependent manner.

3.1 Introduction

Coffee is more than a commodity. Between 50-100 million people are involved in coffee production [170] and more than 100 million people are estimated to consume at least one cup of coffee per day in the USA [171]. In 2015/2016, coffee production was estimated to be 148 million bags (of 60 kg each) while over the same time period, the consumption was estimated to be 151 million [172]. Therefore, the coffee industry needs to meet consumer demand while overcoming multiple challenges.

Coffee production is dramatically affected by several diseases and pests, all directly or indirectly influenced by environmental conditions. The oldest and most relevant disease is Coffee Leaf Rust (CLR) which is caused by *Hemileia vastatrix* Berk. et Br [28, 173]. CLR fungus negatively impacts the production of specialty quality coffee cultivars of *Coffea arabica* L. grown in tropical and subtropical regions of the world [3-4]. This disease reduces the photosynthetic leaf area and yield by 30-80% in the worst cases [4, 28, 87]. CLR directly or indirectly affects the physiological condition of the plant which ultimately translates into cup (tasting) defects and bad coffee. To mitigate the CLR problem, resistant varieties need to be bred with improved market quality profiles. In addition to the lack of genetic knowledge of both plant, disease, and their interaction, the

end use beverage chemistry and sensory perception are other barriers which need to be overcome in the short term.

Polysaccharides constitute 48 to 60% of the dry matter in Arabica coffee beans [77, 88]. The second highest fraction (15-18%) is lipids [88, 174], which are complex and unexplored. The lipid fraction is mainly composed of triacylglycerols (TAG) [175] synthesized from acyl-CoA dependent and independent pathways [176]. Linoleic (C18:2), linolenic (C18:3), oleic (C18:1), and palmitic acids (C16:0) represent between 45%, 3.2%, 8.7%, and 32% of all the fatty acids in the coffee bean, respectively [177]. These fatty acids are highly reduced forms of carbon and among the most important reserves of energy in seeds [176]. Fatty acids are considered of importance as quality attributes due to their carrier function of the different aromatic volatiles [178]. However, the oxidation of unsaturated fatty acids (18:1, 18:2, and 18:3) produces aldehydes during the roasting process which has been considered to reduce acidity, fragrance, body, and flavor of the final beverage [179]. Overall, higher levels of sucrose and lower levels of quinic acid, choline, acetic acid, and fatty acids are markers of good quality [89].

We previously showed that under CLR stress, sensory and chemical volatiles are changed in the coffee beans [180]. We sought to expand that initial study by using fruit samples from the same experiment to determine if the volatile precursors present in the unroasted green beans were associated with some of the chemical volatiles previously detected in roasted beans. Furthermore, from the same bean samples, RNA was obtained in order to evaluate if changes in gene expression could be associated with the volatile profiles and plant stress.

To our knowledge, there is no report on how gene expression varies in response to CLR stress in coffee beans. In order to understand how chemical rust control and high yield production (through fruit thinning) impact the transcriptome in immature and mature coffee beans, the gene expression of beans was compared and analyzed to find candidate genes and/or metabolic pathways that may be related to coffee's response to stress and/or volatile precursors in unroasted green beans. The association between the candidate genes/pathways and the volatiles will serve as a basis to screen these genes for later validation and metabolomics analysis.

3.2 Materials and Methods

3.2.1 Experimental design

The experiment was established as has previously been reported by Echeverria-Beirute, et al. [180] (Chapter II). In summary, the treatments involved two CLR susceptible adult coffee cultivars (*Coffea arabica* L.): an inbred (Red Catuai 44, F_8 originated from 'Caturra' x 'Mundo Novo') and a hybrid (H3, F_1 of 'Caturra' x 'Ethiopian 531'). The cultivars were subjected to manual fruit thinning (0% or 50% removal after pollination) and rust control (with or without cyproconazole and epoxiconazole spray application). The experimental design was a split-split-split plot summarized in Table 4. The control (C) treatment didn't have the rust control (R) and fruit thinning (T) treatments, and represented the most stressful condition in this study.

Table 4. Treatments evaluated in the experiment. Each treatment had four repetitions in the field.

Treatment	Code	Cultivar (G)	Fruit thinning (T)	Rust control (R)
1	R	Inbred	0%	Yes
2	C	Inbred	0%	No
3	R+T	Inbred	50%	Yes
4	T	Inbred	50%	No
5	R	Hybrid	0%	Yes
6	C	Hybrid	0%	No
7	R+T	Hybrid	50%	Yes
8	T	Hybrid	50%	No

3.2.2 Volatile compound analysis

The profiles of volatile compounds from the green beans were obtained as previously reported by Echeverria-Beirute, et al. [180] (Chapter II). A total of 28 key odorants were identified using the Solid Phase Micro Extraction (SPME), Gas Chromatography-Mass Spectrometry (GC-MS) at Texas A&M University (TAMU). The total ion count – area under the curve of each peak (relative abundance) for each volatile was reported and used for later statistical analysis. The relative abundance of the volatile precursors obtained was statistically compared between treatments using an analysis of variance (ANOVA) and *t* test using a cutoff value of $p \leq 0.05$.

A Spearman correlation was performed to compare the volatile precursor's between them and with the rust sporulation (RE) and yield (TH) from each cultivar, according to the values reported in Table 3 [180]. Significance of the tests was determined using a cutoff value of $p \leq 0.05$ and 16 observations per comparison. The statistical analyses were performed in the JMP Pro 13.0.0. (SAS Institute Inc. USA) software package.

3.2.3 RNA isolation

Fruit samples were collected in the experimental plot once during the highest infection phase of CLR disease and fruit harvest (November), between 9:00 and 11:00 a.m. on a drizzle-cool day. Two mature fruits (red color) were manually collected from each plant within a plot that represented a replication and treatment. The fruits were immediately hand-depulped and packed in aluminum envelopes. Each bulked sample of the 10 total plants was immediately placed in liquid nitrogen inside a foam cooler. The same procedure was considered when collecting the immature (yellow-orange color) fruits. Both mature and immature samples were stored in separate envelopes. The samples were later transported inside an insemination tank and stored in a -80°C freezer at the Centro de Investigaciones en Biotecnología (CIB) laboratory of the Instituto Tecnológico de Costa Rica (ITCR, Cartago, Costa Rica) until their use.

In the laboratory, the frozen tissue was quickly ground in a mortar and homogenized in liquid nitrogen. Approximately one hundred milligrams of each ground sample was suspended in the extraction buffer supplied in the PureLink® RNA Mini Kit (LifeTechnologies Inc.) in a 1.6 µl microcentrifuge tube. The extraction of the RNA was performed according to the manufacturer's protocol. In each extraction process, RNA concentration and contamination were analyzed spectrophotometrically at 260 and 280 nm on a DeNovix DS-11 Spectrophotometer. Quality analysis was assessed following electrophoresis on a 1.5% agarose gel.

Dehydration and stabilization of the RNA samples for long term storage and normal temperature transportation were done using the RNAstable® solution

(Biomatrica Inc.). The dehydration followed the manufacturer's protocol. Briefly, 100 µg of the RNA sample was mixed with 20 µl of RNAstable® solution and later slowly dried in a SpeedVac Concentrator (Thermo® Savant DNA 110) for one hour at ambient temperature. The dehydration process was performed in the Laboratorio de Biotecnología en Ciencias Agrarias of the Universidad Nacional (UNA, Heredia, Costa Rica).

3.2.4 cDNA library synthesis

The RNA samples were sent to Polar Genomics LLC (Ithaca, NY). Strand-specific RNA-seq library construction was performed using their own developed protocol compatible with the TrueSeq Stranded Total RNA Library preparation kit (Illumina®), based on Zhong, et al. [181]. All cDNA libraries obtained from each RNA sample treatment were size selected by AMPure XP Beads and then PCR amplified using Illumina primers. The library quality determination was done using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA).

3.2.5 RNA-seq analysis

Sequencing of the libraries was conducted with the Illumina HiSeq2500 (Illumina®), using a single–end, 101 bp strategy at the Institute of Biotechnology at Cornell University. Quality and quantity of the resulting reads were analyzed with the FastQC Software v.0.11.5 [182]. Removal of primer, adapters, and other contaminate sequences from the fastq files were performed using Trimmomatic with default

parameter settings [183]. Twenty-three immature and 23 mature libraries were successfully constructed and sequenced, representing 3 biological replications for 7 of the 8 treatments for each maturity stage. Only treatment 1 from immature fruits and treatment 7 from mature fruits had two biological replications (Table 4) due to failed libraries. The fastq filtered reads were processed using the CLC Genomics Workbench v.9.5.2 software (QIAGEN®, Aarhus, Denmark). RNA-seq analysis was performed in the CLC Genomics Workbench using the annotated *Coffea canephora* genome [134] as the reference genome for annotation. The reads were mapped using the following parameters: mismatch cost of 2, insertion and deletion cost of 3, length fraction of 0.8, similarity fraction of 0.8, global alignment = yes, map to intergenic regions = yes, strand specific = both, maximum number of hits for a read = 10, expression value = total counts and use EM estimation = yes. The output from the RNA-seq analysis tool within the CLC Genomics Workbench was a gene expression (GE) table for each sample containing the counts and expression values for each gene along with its description. The sequencing depth per gene annotated was calculated as reported by Dugas, et al. [136] considering the total number of bases mapped to a gene (exons only) divided by total gene (exon) length. Genes with sequencing depth of 0.5X or higher, were considered ideal for downstream analysis; however, gene expression analysis was performed without filtering sequencing depth, since gene dispersion was adjusted by the negative binomial variance [184] in order to find changes in gene expression.

3.2.6 Gene expression profiles

In order to understand which genes were either differentially expressed or not in response to the maturity stage (immature vs mature), treatments [rust control (R), fruit thinning (T), both (R+T), and no control (C)], and cultivars (inbred vs hybrid), contrast tests were performed according to the desired comparison (Table 6). Contrast tests using the empirical analysis of digital gene expression data in R (edgeR) were performed in the CLC Genomics Workbench v.9.5.2 software. Each gene-annotated value was quantile normalized (Q) as recommended by Bullard, et al. [185] in the CLC Genomics Workbench v.9.5.2 software. In order to control for multiple factors and unbalanced data, the annotated genes were marked as significant differentially expressed genes (DEGs) using a very conservative approach by considering a Bonferroni correction and a false discovery rate (FDR) of $p \leq 0.01$, and a fold change (FC) cut off of log2 greater than or equal to 1 or less than or equal to -1, as limiting conditions. The normalized fold change for each annotated gene in response to the variables (maturity stage, treatments, and cultivars), were clustered in a principal component analysis (PCA) using the restricted maximum likelihood (REML). The estimate of the percentage of variation explained [186] and simplification of the complexity to trends and patterns [187], was used to control the overall error due to multiple sources of variability. Venn Diagrams (Bioinformatics & Evolutionary Genomics, Gent, Belgium, http://bioinformatics.psb.ugent.be/webtools/Venn/) were used to compare and visualize the DEGs according to the experimental conditions to represent shared and unique groups. The DEGs were also used to perform biological function predictions using gene

ontology (GO) analysis with the Singular Enrichment Analysis (SEA) tool in AgriGO v2.0 [188] (http://systemsbiology.cau.edu.cn/agriGOv2/index.php) according to the *Coffea canephora* annotation [134]. Significant GO terms were found using the default FDR $p \leq 0.05$ cutoff value.

3.2.7 Candidate genes associated with chemical volatiles

The DEGs either in immature or mature samples were used to find significant volatiles associated between cultivars using a Spearman correlation ($p \leq 0.05$). The variation and association of the volatiles with the treatments, were represented by a principal component analysis (PCA) using the restricted maximum likelihood (REML). All statistical analyses were performed in the JMP Pro 13.0.0. (SAS Institute Inc. USA) software package.

3.3 Results

3.3.1 Quality volatile precursors

From the 28 chemical volatile compounds detected in the SPME/GC-MS, seven were statistically different when compared between the treatments and/or genotypes (Figure 16, Appendix 4). Overall, the volatiles (E,E)-2,4-Decadienal, (E)-2-Decenal, and nonanal had lower relative abundance under rust control treatments (R); while ethanol and dimethyl sulfide were higher under less stressed treatments (R, T, or R+T). Acetic acid and dl-limonene were only detected in the hybrid. The origin and expected impact on cup quality are summarized in Table 5.

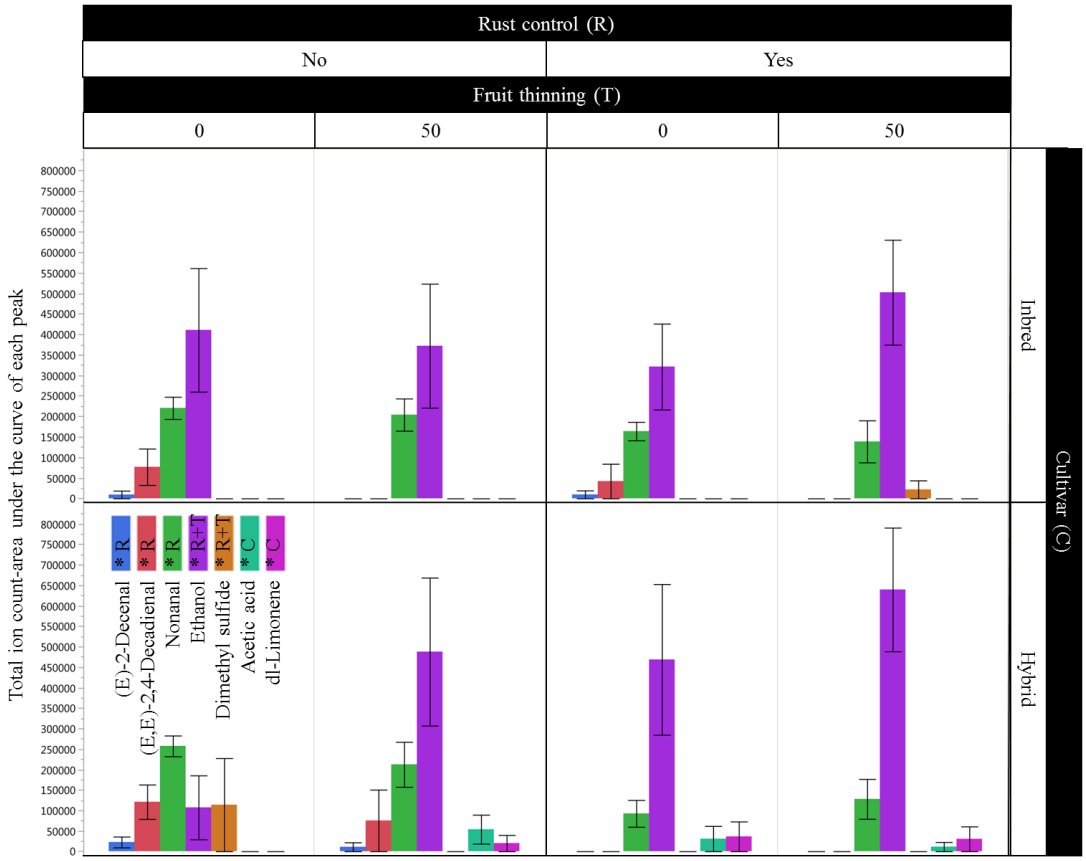


Figure 16. Relative abundance of significant volatile precursors found to be affected by the treatments.

Error bars represent the standard error (SE). Legend below the top left panel shows the identity of the volatile and the significant difference (*, $p<0.05$) found between rust control (R), thinning (T) or rust plus thinning (R+T) treatments, or cultivars (C1) using a *t* test. The volatiles (E,E)-2,4-Decadienal, (E)-2-Decenal, and nonanal were significantly lower under rust control treatments (R). Ethanol and dimethyl sulfide were higher under rust control and fruit thinning (R+T) treatments. Acetic acid and dl-limonene were only detected in the hybrid, especially under less stressful conditions (Appendix 4).

Table 5. Volatile precursor's origin and expected impact on beverage quality after roasting.

Volatile precursor	CAS Registry Number [189]	Origin	Compound formed after roasting	Impact odor
(E,E)-2,4-Decadienal	25152-84-5	Autooxidation of linoleic acid [190]	Aldehydes, furans	Fatty, fried soapy, green [190]
(E)-Decenal	3913-81-3	Autooxidation of oleic acid [190]	Aldehydes, furans	Pungent, green [190]
Nonanal	124-19-6	Autooxidation of oleic or linoleic acid [190]	Aldehydes, furans	Pungent, green
Ethanol	64-17-5	Pyruvate or sugar fermentation	Furans and other alcohols [191]	Alcoholic, medicinal [192]
Dimethyl sulfide	75-18-3	Sulfuric compounds (SAM, cysteine, glutathione, or methionine) [193]		Sulphur, cabbage [189]
Acetic acid	64-19-7	Carbohydrate degradation (arabinose from arabinogalactans) [194]	Pyridines by pH acidification	Acid[194], vinegar
dl-limonene	138-86-3	Geranyl diphosphate [195]	Limonene	Flowery note [196]

3.3.2 Volatile precursors' correlation with rust sporulation and yield

Considering the volatiles, rust sporulation (RE), and yield (TH) from Table 3 [180] (Chapter II), the Spearman analysis showed significant correlations between the cultivars (Table 6). The inbred only showed a positive correlation of the (E,E)-2,4-Decadienal volatile as the yield (TH) increased. The same previous volatile, nonanal, and (E)-2-Decenal, also showed a positive increase with increase of rust sporulation (RE) in the hybrid. The ethanol volatile however, decreased with increase of RE, (E,E)-2,4-Decadienal, nonanal, and (E)-2-Decenal in the hybrid. Acetic acid and dl-Limonene were only found in the hybrid and correlated with ethanol and nonanal, respectively,

suggesting metabolic pathway variations in beverage quality according to fruit load and rust epidemics for this particular cultivar.

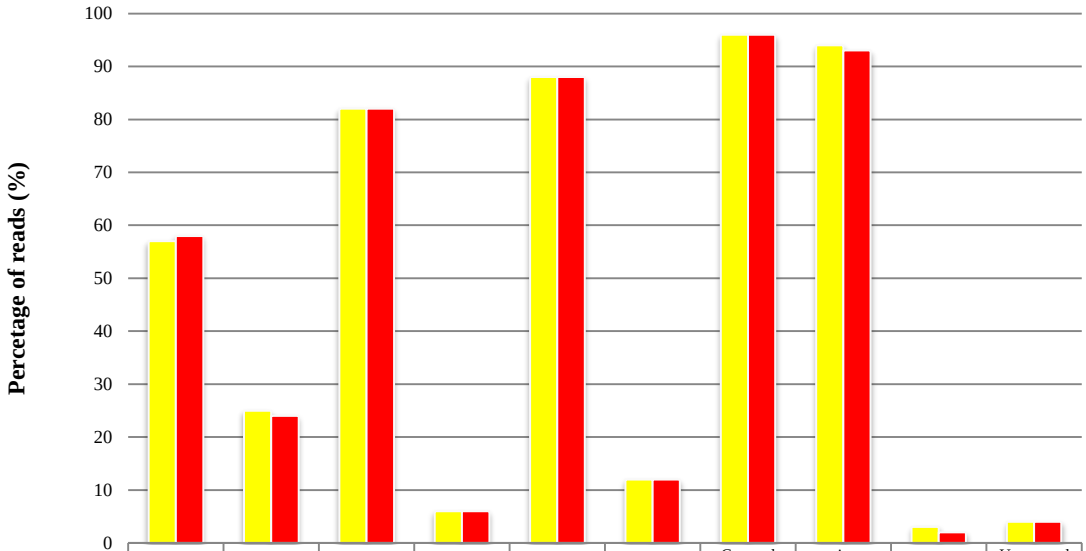
Table 6. Volatile precursor correlations within them, rust sporulation (RE), and yield (TH) according to the cultivar.

Cultivar	Volatile precursor	Trait/Volatile	Spearman ρ	*p*-value
Inbred	(E,E)-2,4-Decadienal	TH	0.570	0.021
Hybrid	(E,E)-2,4-Decadienal	RE	0.574	0.020
Hybrid	Ethanol	RE	-0.646	0.007
Hybrid	Nonanal	RE	0.731	0.001
Hybrid	(E)-2-Decenal	RE	0.636	0.008
Hybrid	(E,E)-2,4-Decadienal	Ethanol	-0.639	0.008
Hybrid	Nonanal	Ethanol	-0.710	0.002
Hybrid	(E)-2-Decenal	Ethanol	-0.636	0.008
Hybrid	Nonanal	(E,E)-2,4-Decadienal	0.678	0.004
Hybrid	(E)-2-Decenal	(E,E)-2,4-Decadienal	0.879	0.000
Hybrid	(E)-2-Decenal	Nonanal	0.708	0.002
Hybrid	Acetic acid	Ethanol	0.677	0.004
Hybrid	dl-Limonene	Nonanal	-0.499	0.049

3.3.3 Fruit transcriptome

A total of 1,172,573,476 high-quality sequences were obtained from the 46 RNA samples. Approximately 97% of the sequences were between 100-101 bp length after trimming, with 44.5% GC content, and a phred score showing that 98% of the sequences were higher than 30 (99.9% accuracy in base calling) (Appendix 5). On average, 82% of the sequences aligned to exons and over 93% of the fragments were uniquely mapped to the diploid *C. canephora* [134] reference genome (Figure 17, Appendix 6). Across maturity stages and treatments, overall gene expression with a sequencing depth higher than 0.5X

was observed for 4,289 annotated genes (19% of the average 23,230 unigenes) (Appendix 7).



	Exon	Exon-exon	Total exon	Intron	Total gene	Intergenic	Counted fragments	unique fragments	- non-specifically	Uncounted fragments
Immature (%)	57	25	82	6	88	12	96	94	3	4
Mature (%)	58	24	82	6	88	12	96	93	2	4

Figure 17. Immature and mature bean distribution of the total number of all sequencing reads that passed Illumina's filtering among annotated features across the *Coffea canephora* genome.
More details in Appendix 6.

3.3.4 Contrasts with the largest effect differences

The principal component analysis (PCA) of the expression data (Figure 18) indicated that gene expression exhibited increasingly larger variation between 1) different treatments of a single cultivar and maturity stage, 2) maturity stages within cultivars, and 3) cultivars. The proportion of the treatment effect was higher when compared to the cultivar effect, corresponding to our previous findings of phenotypic data (Figure 9) as well, so we have organized the results to follow these contrasts using

the maturity stage as a conditional factor. We follow up with analysis of important DEGs across the comparisons.

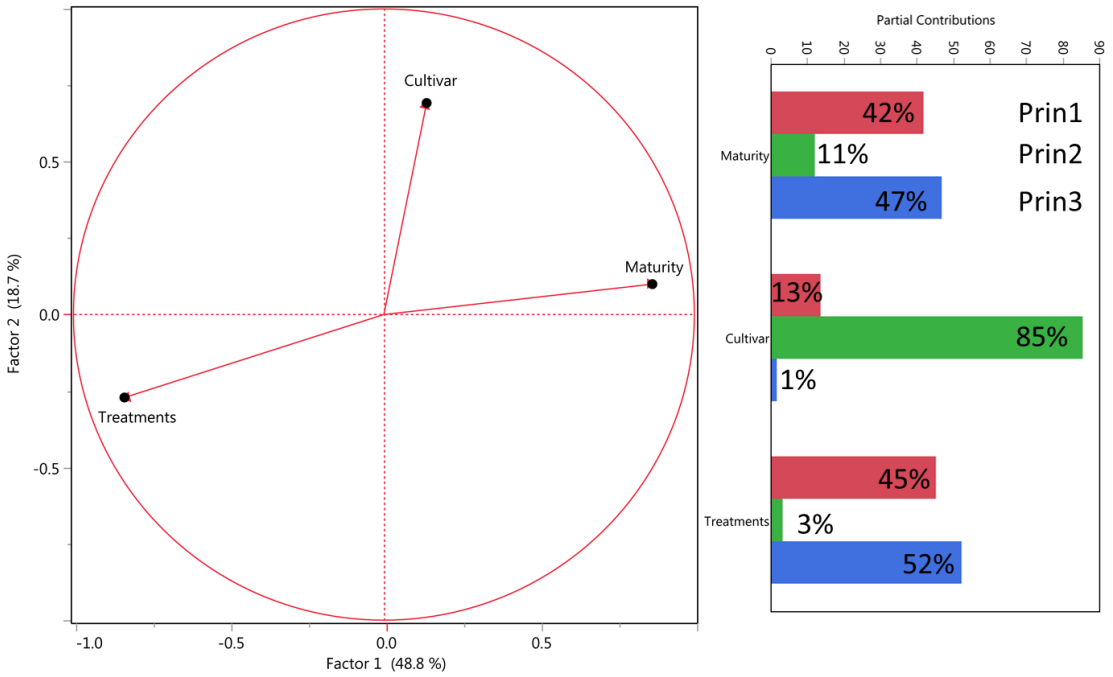


Figure 18. Principal component analysis (PCA) representing the effects of the sources of variability maturity, treatment, and cultivar using all significant annotated genes in this study.
The first two factors accounted for the 67.5% of the variability of the gene expression. The partial contribution for the first and largest principal component (red colored bar) showed that 42%, 13%, and 45% of the variability was represented by the maturity stage, cultivars, and treatments, respectively; while the cultivar effect was the highest in the second principal component (green colored bar) with 85% of the total variance.

3.3.5 Comparing immature samples versus mature samples

All the genes significantly expressed across the maturity stages of both cultivars and all treatments, revealed that between 38% and 36% of the annotated genes (8,893 in the immature and 8,458 in the mature from the average of 23,230 unigenes) were expressed with at least a |2| fold change or higher between maturity stages. Comparing

all samples using maturity stage as a predictor, a total of 471 DEGs were found to be statistically significant (Table 7, Comparison #1, Appendix 8). Ninety percent (426) of the 471 DEGs corresponded to higher expression in the immature samples, showing that most metabolic processes were active before ripening. The immature stage was selected to find treatment and cultivar significant differences of gene expression.

Table 7. Summary of the fruit transcriptome comparisons evaluated in the experiment.

Comparison	Maturity stage	Treatment	Cultivar	Results
1	Immature vs Mature	All	Both	471 DEGs (426 up-regulated in the immature and 45 in the mature)*
2	Immature	Each treatment	Both	309 DEGs shared between cultivars. Treatment DEGs as follows: 51 in C, 91 in R, 96 in T, and 71 in R+T *
3	Immature	All	Inbred vs Hybrid	208 DEGs unique (180 upregulated in the hybrid, 28 upregulated in the inbred)*

* normalized fold change >|2|, Bonferroni correction *p*-value <0.01, FDR *p*-value < 0.01.

3.3.6 Comparing treatment effects by cultivar

In total 309 genes were differentially expressed across treatments within specific cultivars (Table 7, Comparison #2; Figure 19). The analysis showed that less than 49% (208) of those DEGs in immature beans were also found to be significant when considering the cultivar. These represent the core genes that were affected by environmental changes. The number of differentially expressed genes according to each treatment showed that the hybrid had between 6 to 7 times more DEGs than the inbred, 66% (271) of those DEGs were unique between treatments (Figure 19). In the inbred,

74% of the 38 shared DEGs were unique to a treatment, showing both fewer DEGs expressed and fewer DEGs shared than in the hybrid.

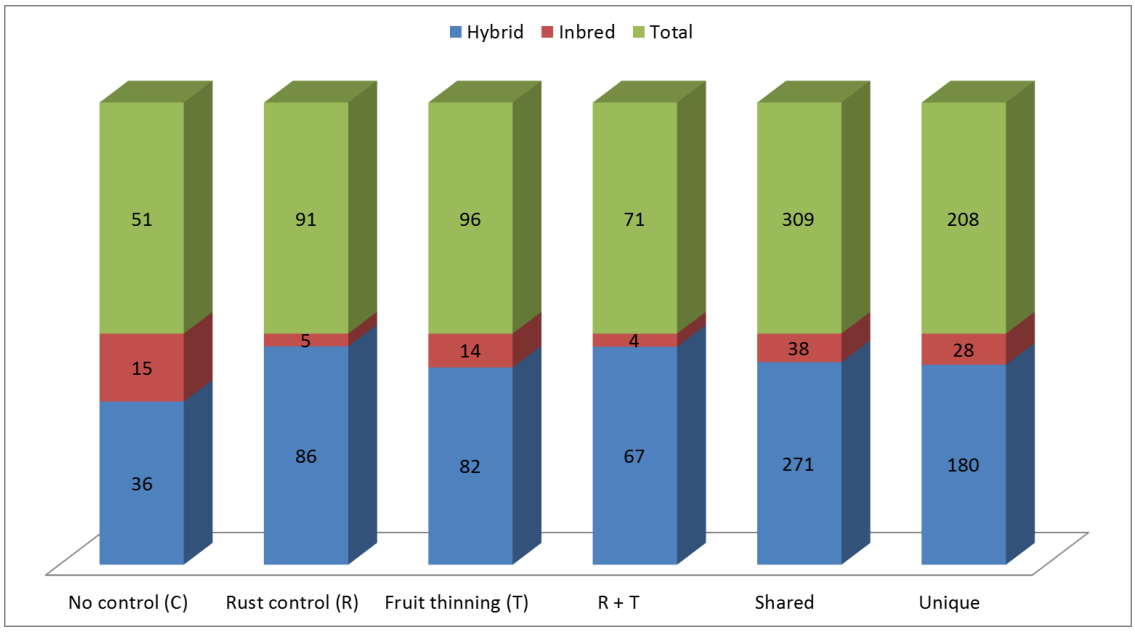


Figure 19. Number of differentially expressed genes (DEGs) between treatments in the immature samples.
Shared DEGs for control (C), rust control (R), fruit thinning (T), and R+T treatments are shown for each cultivar to exemplify the proportions of DEGs within the immature stage. Shared DEGs represent total number of DEGs between treatments (sum of all C, R, T and R+T contrasts). Unique DEGs represent the number of DEGs found, regardless of how many times it was present in other treatments.

3.3.7 Comparing cultivars

Comparing the inbred to the hybrid cultivar, 208 genes were differentially expressed (Table 7, Comparison #3; Appendix 9; Figure 19 - Unique) with 180 genes up-regulated in the hybrid and the remaining 28 genes up-regulated in the inbred in the immature beans (Figure 20, Appendix 9).

The most shared DEGs between treatments (1 in Figure 20A; 3 in Figure 20B) were found for each cultivar. Just one DEG annotated as a respiratory burst oxidase

homolog protein B (RBOH) was found to be involved in all treatment interactions in the inbred (Table 8, Gene # 1), either highly expressed under treatments or control. In the hybrid, three DEGs annotated as a cytochrome P450 86A2, potassium transporter 5, and 3-ketoacyl-CoA synthase 17, were involved in all treatment interactions (Table 8, Genes # 2-4). Multiple DEGs related to the treatments can represent higher levels of gene networking, spatial–temporal organization, or amount of alternative pathways related as a response to stimuli.

A. Inbred's DEGs by treatment **B. Hybrid's DEGs by treatment**

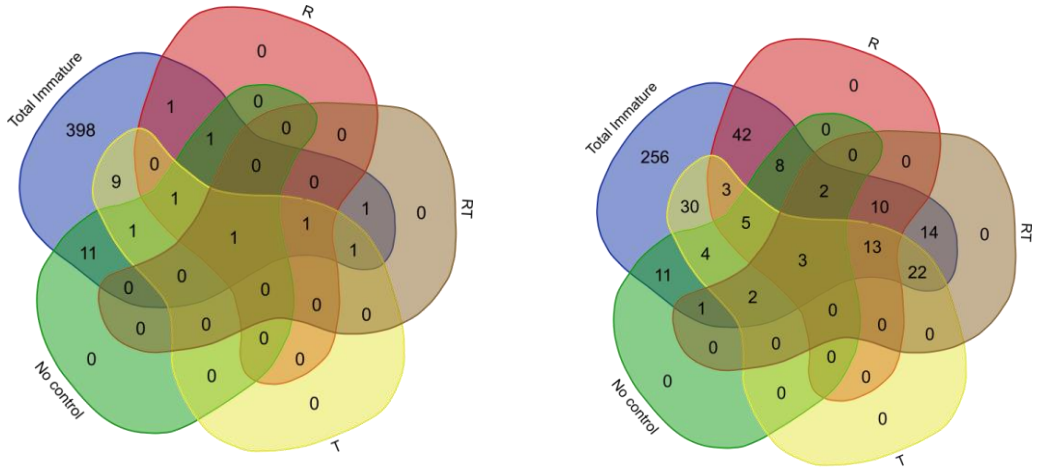


Figure 20. Venn diagrams comparing the total DEGs between cultivars according to the treatments.
Shared and unique DEGs by treatment are represented for the inbred (A) and the hybrid (B) (Appendix 9). Each color represents a treatment. The numbers represent the shared or unique DEGs obtained directly after the edgeR analysis.

Table 8. Relevant DEGs found in different interactions.

Gene #	Maturity stage	Treatment	Cultivar	Annotated gene (Gene ID[2])	Function	Reference
1	Immature	All	Inbred	respiratory burst oxidase homolog protein B (RBOH) (Cc02_g33130)	key role in the network of ROS production	[197]
2	Immature	All	Hybrid	cytochrome P450 86A2 (Cc02_g38150)	hydrolysis of fatty acids, synthesis of phytooxylipins or other protective biopolymers such as cutin and suberin	[198]
3	Immature	All	Hybrid	potassium transporter 5 (Cc05_g03840)	osmotic/drought stress	[199]
4	Immature	All	Hybrid	3-ketoacyl-CoA synthase 17 (Cc03_g04570)	fatty acid biosynthesis under osmotic stress	[200]

[2] Gene ID is from the *C. canephora* gff file that can be downloaded from 134.Denoeud, F.; Carretero-Paulet, L.; Dereeper, A.; Droc, G.; Guyot, R.; Pietrella, M.; Zheng, C.; Alberti, A.; Anthony, F.; Aprea, G.; Aury, J.-M.; Bento, P.; Bernard, M.; Bocs, S.; Campa, C.; Cenci, A.; Combes, M.-C.; Crouzillat, D.; Da Silva, C.; Daddiego, L.; De Bellis, F.; Dussert, S.; Garsmeur, O.; Gayraud, T.; Guignon, V.; Jahn, K.; Jamilloux, V.; Joët, T.; Labadie, K.; Lan, T.; Leclercq, J.; Lepelley, M.; Leroy, T.; Li, L.-T.; Librado, P.; Lopez, L.; Muñoz, A.; Noel, B.; Pallavicini, A.; Perrotta, G.; Poncet, V.; Pot, D.; Priyono; Rigoreau, M.; Rouard, M.; Rozas, J.; Tranchant-Dubreuil, C.; VanBuren, R.; Zhang, Q.; Andrade, A. C.; Argout, X.; Bertrand, B.; de Kochko, A.; Graziosi, G.; Henry, R. J.; Jayarama; Ming, R.; Nagai, C.; Rounsley, S.; Sankoff, D.; Giuliano, G.; Albert, V. A.; Wincker, P.; Lashermes, P., The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* **2014,** *345* (6201), 1181-1184.

Table 8 continued…

Gene #	Maturity stage	Treatment	Cultivar	Annotated gene (Gene ID[2])	Function	Reference
5	Mature	Control	Inbred and Hybrid	putative 14 kDa proline-rich protein DC2.15 (Cc00_g00780 & Cc00_g00800)	cell wall modification and organization	[201]
6	Mature	Control	Inbred and Hybrid	putative disease resistance-responsive (dirigent-like protein) family protein (Cc00_g04320)	disease defense early response	[202]
7	Mature	Control	Inbred and Hybrid	GDSL esterase/lipase At1g71250 (Cc10_g16390)	regulates plant development, morphogenesis, synthesis of secondary metabolites, and defense responses	[203]
8	Immature	Control, R+T	Inbred, Hybrid	endoglucanase 11 (Cc05_g03130)	cleaves internal β-1,4-glucosidic bonds	[201]
9	Immature	R	Hybrid	xyloglucan endotransglucosylase/hydrolase proteins (XTH) (Cc06_g13440)	increasing the cellulose and xyloglucans link	[204]
10	Immature	R	Hybrid	Pectinesterase (Cc06_g04690)	esterification of pectin into carboxylate groups	[205]
12	Mature	All	Inbred & Hybrid	8-hydroxyquercetin 8-O-methyltransferase (Cc00_g12970)	enlarge the backbones of rich nitrogen-containing compounds	[206]
13	Immature	Control	Hybrid	aluminum-activated malate transporters (Cc00_g23280)	regulate the malate accumulation and release from the vacuole	[207]
14	Immature	Control	Hybrid	mitochondrial 2-oxoglutarate/malate carrier protein (Cc07_g14260)	nitrogen assimilation and carrier	[208]

3.3.8 Gene ontology (GO) analysis

Analysis of the GO terms within both cultivars and specific to maturity stage (Table 7, Comparison #1), revealed that different processes were activated when comparing the immature and mature samples (Appendix 10 & 11). The immature samples were especially enriched in fatty acid biosynthetic process (GO:0006631), carbohydrate metabolic process (GO:0005975), Redox (GO:0055114), cell wall components (GO:0005618), and molecular functions related to sugar binding (GO:0005529), iron ion binding (GO:0005506), oxidoreductase activity (GO:0016491), xyloglucan:xyloglucosyl transferase (GO:0016762), and aspartic-type of endopeptidase activity (GO:0004190). The mature samples were enriched only in molecular functions (GO:0003674) specifically related to *O*-methyltransferase activity (GO:008171) and protein dimerization (GO:0046983). The overall gene expression showed higher activity in the immature stage than in the mature stage (ripening).

3.3.9 Volatile precursors correlated with the DEGs

The Spearman correlation showed that 16 DEGs (3% of the 471 mature or immature significant DEGs) were significantly associated ($p \leq 0.05$) with the volatiles according to the maturity stage or cultivars across all treatments (Appendix 12).

In a PCA that combined DEGs with volatiles, the first two factors represented 78% of the inbred variability and 99% of the hybrid variability (Figure 21). In both cultivars, similar clustering of volatiles (E)-Decenal and (E,E)-2,4-Decadienal was observed with DEGs comprising a putative 14 kDa proline-rich protein DC2.15 (Proline-

rich 1 and 2) and a putative disease resistance-responsive (dirigent-like protein) family protein (Dirigent-like protein) (Table 8, Genes #5 and #6). Both of these DEGs increased their normalized expression under control treatments (C). The volatile nonanal was also correlated with the GDSL esterase/lipase At1g71250 (GDSL esterase/lipase) under control treatments (C) in both cultivars (Table 8, Gene #7). The volatile dl-limonene showed higher correlations with eleven of the total DEGs, but only in the hybrid under rust control treatment (R).

As reported before in Table 6, the same clusters of correlated variables between the volatiles (E,E)-2,4-Decadienal, nonanal, and (E)-2-Decenal with the increase of rust sporulation (RE) in the hybrid, are also present in Figure 21. The increase of rust sporulation (RE) under no fruit thinning and no rust control treatment, suggests being the main reason of change in gene expression in this cultivar.

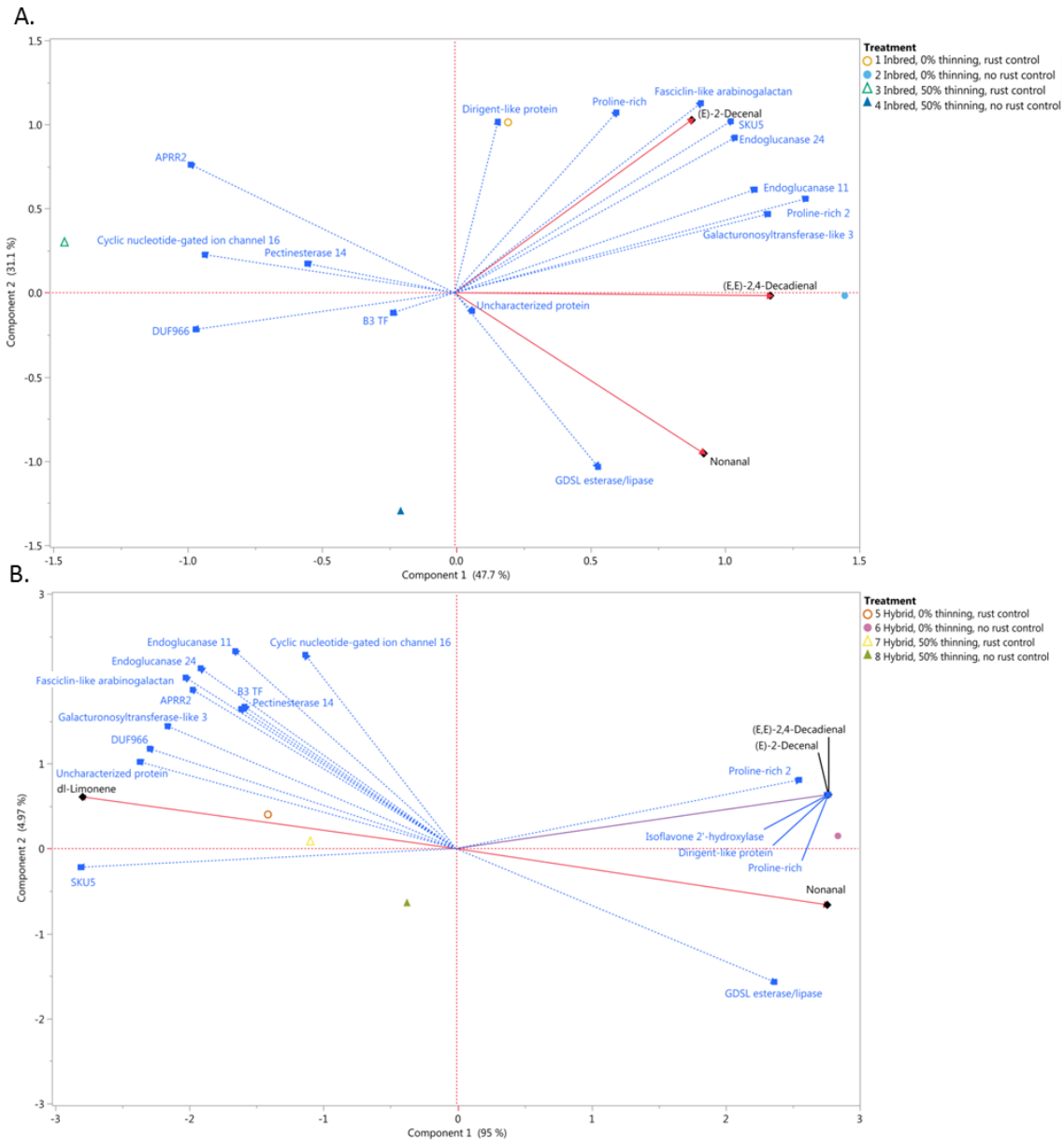


Figure 21. Principal component analysis (PCA) of the correlated DEGs and volatile precursors.
The treatments for the inbred A) and hybrid B) show that a higher variance was found under more stress (i.e. control treatments at the right of the biplot graph).

3.4 Discussion

3.4.1 Fruit transcriptome

The transcriptome of the coffee fruit in two different cultivars and at two maturities was significantly influenced by two types of management practices intended to alleviate coffee leaf rust disease (CLR). The transcriptome and functional annotation of a total of 471 differentially expressed genes (DEGs) into 19 gene ontology (GO) terms, reflects that the coffee bean gene expression and volatile profiles, can be significantly modified by management practices.

Our statistical analysis used a conservative approach in the multiple comparisons that were analyzed. We used both the Bonferroni correction which controls the familywise error rate, and the false discovery rate (FDR) which controls the false positive rate [209]. Even when a large number of genes were marked as not statistically significant at 99% confidence, we still found many DEGs that were consistent across the edgeR comparisons.

Similar to our previously reported phenotypic data (Figure 9) [180], the treatment effect explained more variance (more than 20%) than the cultivar effect (8%). However, since the cultivar effect was the highest for the second principal component (85%), we considered the maturity stage as a conditional variable in later analysis. The control of the maturity stage, allowed the comparison of the treatment effect in the cultivar to find differentially overexpressed genes (DEGs) related to management practices.

3.4.2 DEGs involved in functional gene ontologies (GO)

Of all significant DEGs, 90% were found in the immature bean (Table 7). The higher expression of genes in the immature bean was related to lipid metabolism, cell wall, carbohydrate metabolism, and oxidation-reduction. During the immature stage, it is expected that an active synthesis of sources of energy and accumulation in the endosperm in the form of triacyl glycerol (TAG) or chlorogenic acids (CGA) occurs [77], which were in accordance with our findings.

One of the DEGs that was expressed in both cultivars was an endoglucanase (Table 8, Gene #8), a cellulase type of enzyme that cleaves internal β-1,4-glucosidic bonds and is classified into the carbohydrate metabolism process (GO:0005975). The endoglucanase in our study was highly expressed in the inbred under the control treatment, but was high in the hybrid under rust control or fruit thinning treatments (Appendix 9). The difference between both cultivars in carbohydrate metabolism process can be related to regulation at the transcription level or coordination of enzymes such as xyloglucan endotransglucosylase/hydrolase proteins (XTH).

The XTH (Table 8, Gene #9) jointly with the expansins, loosens the cell wall in order to increase its growth [204]. XTH increased with rust control treatments in the hybrid during the immature stage. The evidence suggests that under less stress (such as with rust control), the hybrid continued cell wall expansion, but its activity decreased under higher stress. Bote and Jan [153] reported an increase of 11-14% in bean diameter when coffee plants were fruit thinned at 75%, showing that under less fruit load, changes in bean size can be expected. Since the coordination of both endoglucanases and XTH are responsible for the expansion of the cell walls, our observations suggests that the hybrid

had better molecular response coordination at a cell wall level in order to defend itself by reducing cell wall expansion during stress.

Part of the network interactions of cell wall structure, stress, and quality, is also related to pectinesterase activity (Table 8, Gene #10). The pectinesterases are responsible for the esterification of pectin into carboxylate groups helping calcium mediated responses. The calcium mediated responses reduce the cell wall matrix while recruiting jasmonic and salicylic acids [205]. The pectin breakdown can indirectly lead to ROS, kinase cascades, and pathogenesis-related genes [210-211]. In our study, pectinesterase was higher in the hybrid under rust control treatments (R and R+T). We hypothesize that the interaction between pectinesterases, expansins, endoglucanases, and XTH are associated with an oxidative burst which controls the peroxidation of the cell wall fatty acids under stress.

The ROS activity network seems to be the reason why the two cultivars had different defense responses. One respiratory burst oxidase homolog protein B (RBOH; Table 8, Gene #1) out of six found in the *C. canephora* annotation database (not shown), was the only DEG varying its expression in multiple interactions in the inbred and it is associated with early stress response [197]. However, the hybrid had three DEGs differentially expressed between all treatments: a) the cytochrome P450 86A2, related to the hydrolysis of fatty acids in order to synthesize phytooxylipins or other protective biopolymers such as cutin and suberin [198]; b) the potassium transporter 5, related with osmotic/drought stress [199]; and c) 3-ketoacyl-CoA synthase 17, related to fatty acid biosynthesis under osmotic stress [200]. These three DEGs are associated with a later

response to abiotic/biotic stress. Since the balance of fatty acid peroxidation can enhance signal amplification and defense responses by synthesis of phytooxylipins and other volatiles [212], we suggest that the hybrid had an earlier and enhanced defense response, while the inbred was just starting.

3.4.3 Volatile precursors associated with the DEGs

Evidence of a direct host-pathogen interaction was found from two DEGs. The correlation of proline-rich proteins and the putative disease resistance-responsive (dirigent-like protein) family protein with the volatiles (E)-Decenal and (E,E)-2,4-Decadienal (Figure 21), suggests CLR haustorial penetration activity. Proline and glycine-rich proteins are associated with a collection of fragmented parts of the cell wall, which may serve as signal peptide for development or abiotic stresses [213-214]. The dirigent-like protein was previously associated with an early response to infection in coffee plants susceptible to CLR [201]. In our study, higher levels of (E,E)-2,4-Decadienal, (E)-2-Decenal, and nonanal, were found under the control treatment, which had higher rust incidence (12%) and rust sporulation (27%) as we previously reported [180]. Both candidate genes correlated to fatty acid peroxidation, suggesting that the cell wall of the fruits were facing active structural changes due to biotic stress.

The correlation of nonanal with Gly-Asp-Ser-Leu (GDSL) esterase/lipase [215] also appears to be related to biotic stress (Figure 21, Appendix 12). These type of enzymes may regulate plant development, morphogenesis, synthesis of secondary metabolites, and defense responses [202]. In our study, the association with the control treatment (C)

coincides with the increase of susceptibility to CLR. Moreover, nonanal correlation with rust sporulation (RE) (Table 6), suggests synthesis activation from host-pathogen interactions. However, nonanal has been shown to have anti-fungal activity [216], which is expected to be the reason why this compound is synthesized in the plant, even when its effectiveness can be compromised. As a host-pathogen interaction, it is hypothesized that either the GDSL esterase/lipase or the phytooxylipins can be used as markers for CLR pathogenesis [217].

The type and concentration of fatty acids are relevant for cup quality, but have origins in the structural and protective biological functions for the plant. Specifically in coffee, the increase of palmitic and oleic acid have been correlated with higher field temperatures, while linoleic and linolenic acids increase under lower field temperatures [2, 218]. Increased levels of linoleic and linolenic acids have been shown to be related to resistance against *Colletotrichum, Pseudomonas,* and *Botrytis* in avocado, tomato, and beans, respectively [219]. Jasmonic acid (JA), derived from linolenic acid precursors are known to be involved in abiotic and biotic stresses as a hormone regulator [219]. Furthermore, the decrease of the monosaturated oleic acid is also associated with defense signals, up-regulating *R* genes mediated by salicylic acid (SA). Therefore, the variation of fatty acid content and composition in the cell membranes is associated with environmental adaptations to abiotic and biotic stresses.

The defensive relevance of fatty acids is also related to the capacity to mediate responses induced by reactive oxygen species (ROS) [219]. Plant stress responses to abiotic or biotic conditions are normally mediated by the production of ROS which are

controlled in the plant by reduction and oxidation processes (Redox) [220]. The balance and dynamics of Redox interactions generate different signaling networks depending on stress type, duration of the stimulus, genotype, phenology, tissue, etc.

In coffee, ROS activated signals have been detected in CLR compatible (qualitative resistance, host, or gene for gene) and incompatible (quantitative resistance or non-host) interactions [38, 221]. The metabolic pathways related to the balance of ROS impact the effective growth and development which represents sources of genes to target in a plant breeding program focused on stress resistance and quality. Since ROS signals are part of the response to stress and can induce metabolic changes in any part of the plant [197], both plant physiology and bean quality can be products of a defense process.

The volatile dl-limonene was associated with rust control or fruit thinning treatments in our study (Figure 21) and inversely correlated with nonanal (Table 6). Production of dl-limonene can be enhanced by increased carbon flux of glycerol-3-phosphate and pyruvate from photosynthesis metabolism; which are more efficient under lower demands of carbohydrates due to a vegetative and reproductive balance [222]. In fact, the carbohydrate metabolic process (GO:0005975) was the only enriched GO function affected significantly by the fruit thinning treatments in the hybrid. The increase in carbohydrate related DEGs in the hybrid under fruit thinning, seems to be the reason why dl-limonene was found in higher abundance under reduced stress conditions.

Overall, this transcriptomic information can be used to dissect pathways related to biotic stress response and bean chemistry. More importantly, the comprehensive relationships are important to understand in order to improve durable resistance without

decreasing beverage quality. Further studies can target the reported pathways identified in this study using when available, the *Coffea arabica* L. genome reference, in order to validate and expand the list of candidate genes and functions.

CHAPTER IV

CANDIDATE GENES IN COFFEE (*COFFEA ARABICA* L.) LEAVES ASSOCIATED WITH RUST (*HEMILEIA VASTATRIX* BERK. & BR) STRESS

Coffee leaf rust (CLR) caused by *Hemileia vastatrix* Berk. et Br., is one of the most threatening diseases for *Coffea arabica* L. Efforts around the world to breed resistant varieties rely on qualitative and quantitative resistance, most with unknown durability due to a high mutation rate of CLR and the large number of physiological races of this fungus. Under the hypothesis that gene-for-gene pathogen-host interactions are easier to be suppressed, most tolerance to CLR relies on unknown horizontal resistance. This study evaluated gene expression in leaves of two susceptible coffee cultivars under different rust control and fruit thinning treatments. RNA-seq analysis focused on the association of differentially expressed genes (DEGs) related to CLR under different treatments to modify the disease and plant fruit load. Gene expression and gene ontology (GO) analysis allowed the identification of a total of 100 genes associated with quantitative traits. From these, 88 were correlated with rust incidence, rust severity, and rust sporulation. The expression of genes coding for pathogenesis-related proteins was found to increase with more disease in the inbred, while genes involved in homoeostasis and broader cell wall structuring processes were upregulated in the hybrid. The enriched gene functions and associations revealed that a possible hypersensitive response (HR) in the inbred and a systemic acquired resistance (SAR) in the hybrid were involved in the tolerance mechanisms to biotic stress. **Introduction**

Coffee production is dramatically affected by several diseases and pests. The oldest and most relevant disease that has a significant effect on the leaves of susceptible cultivars is coffee leaf rust (CLR) caused by *Hemileia vastatrix* Berk. et Br. [28, 173]. Because CLR coevolved with coffee in Africa, it has been able to development 53 (or more) different physiological races according to the resistance genes developed by the plant [12]. The complex host-pathogen compatible interactions, have been described by a gene-for-gene model [41] governed by nine major *R* alleles (S_H1-S_H9), and other minor mostly-unknown genes which confer partial resistance [27, 223]. The major goal of the breeding programs around the world is to achieve durable combinations of *R* genes with other minor genes in order to prevent the compatible reaction of the CLR races with the host.

Plant stress responses are normally mediated by the production of reactive oxygen species (ROS) which are controlled in the plant by reduction and oxidation processes (Redox) [220]. The interaction of Redox with other molecules such as lipid derivatives, plant hormones like ethylene, jasmonic acid (JA), and salicylic acid (SA), together with nitric oxide (NO), regulate plant homeostasis and morphogenesis [224].

In coffee, ROS activated signals have been detected in CLR compatible (qualitative resistance, host, or gene-for-gene) and incompatible (quantitative resistance or non-host) interactions [38, 221]. However, the identification of the genes involved in the hypersensitive response (HR) or systemic acquired resistance (SAR) in coffee tolerance, remain unknown.

Abiotic and biotic stresses have different signatures and responses. Biotic stresses have been shown to be also modulated by leucine rich repeats (LRR) receptor-like serine/threonine-protein kinases involved in the early signal perception of pathogen effectors [225-226]. Homologues of LRR type of kinases in different plant species, has been identified as resistant genes analogues (RGAs)[57]. The molecular network involved with RGAs transduce signals to plant hormones and histones to enhance signal perception and induce transcription of resistance genes, which over the long term, induces a systemic acquired resistance (SAR) in resistant cultivars.

Our previous findings analyzing changes in the fruit transcriptome showed that the management of the plant altered gene expression in a genotype-dependent manner and that different CLR control treatments and maturity stages were also important. In this study we used the leaves from the same plants to identify profiles of: A) differentially expressed genes attributable to each cultivar and treatment, B) functional gene ontologies enriched under rust control treatments, and C) associate the gene expression and functional categories with the phenotypic data, to suggest candidate genes for future validation studies.

4.2 Materials and Methods

4.2.1 Experimental design

The experiment was stablished as previously reported in Chapter II [180]. The treatments involved two CLR susceptible adult coffee cultivars (*Coffea arabica* L.): an

inbred (Red Catuai 44, F_8 originated from ‘Caturra’ x ‘Mundo Novo’) and a hybrid (H3, F_1 of ‘Caturra’ x ‘Ethiopian 531’). The cultivars were subjected to fruit thinning (0% or 50% after self-pollination) and rust control (with or without cyproconazole and epoxiconazole spray application). The experimental design was a split-split-split plot summarized in Table 3 (Chapter II). The control (C) treatment didn’t have the rust control (R) and fruit thinning (T) treatments, and represented the most stressful condition to the plants in this study.

4.2.2 RNA isolation and transcriptome analysis

Two young leaf samples were collected from each plant and bulked according to their repetition and treatment. Leaf sampling was done in the experimental plot once during the highest infection phase of rust disease and fruit harvest (November), between 9:00 and 11:00 a.m. of a drizzle-cool day. Each bulked sample from 10 total plants (20 leaves) was immediately placed in liquid nitrogen inside a foam cooler. The samples were later transported inside an insemination tank and stored in a -80°C freezer at the Centro de Investigaciones en Biotecnología (CIB) laboratory of the Instituto Tecnológico de Costa Rica (ITCR, Cartago, Costa Rica) until their use.

RNA extraction, quality analysis, and dehydration, was performed as reported in Chapter III. The cDNA library synthesis and sequencing was done using the Illumina HiSeq2500 (Illumina®) using a single-end 101 bp strategy, at the Institute of Biotechnology at Cornell University by Polar Genomics LLC (Ithaca, NY). The quality, removal of primer, adapters, and contaminants, were done Trimmomatic with default

parameter settings [183]. Following cDNA library synthesis and sequencing, we obtained high quality sequence reads for 23 samples, representing almost three biological repetitions for each treatment. The only exception with two repetitions was treatment 7 (Table 3).

RNA-seq analysis was performed using the CLC Genomics Workbench software v.9.5.2 (QIAGEN®, Aarhus, Denmark) using the *Coffea canephora* genome [134] as the reference genome for mapping. The reads were mapped using the following parameters: mismatch cost of 2, insertion and deletion cost of 3, length fraction of 0.8, similarity fraction of 0.8, global alignment = yes, map to intergenic regions = yes, strand specific = both, maximum number of hits for a read = 10, expression value = total counts and use EM estimation = yes. Following mapping of the reads to the annotated *C. canephora* genome [134], the resulting gene expression (GE) annotation table was used for further analysis. The sequencing depth per gene annotated was calculated as followed by Dugas, et al. [136] considering the total number of bases mapped to a gene (exons only) divided by total gene (exon) length. Genes with sequencing depth of 0.5X or higher, were considered ideal for downstream analysis; however, gene expression analysis was performed without filtering sequencing depth, since gene dispersion was adjusted by the negative binomial variance [184] in order to find changes in gene expression.

4.2.3 Differentially expressed genes (DEGs)

We evaluated the quantile normalized (Q) annotated data set in order to understand which genes were differentially expressed and by which cultivar and

treatments [185] as recommended by Bullard, et al. [185] in the CLC Genomics Workbench v.9.5.2 software. The comparison of DEGs within the same treatment but between cultivars was done in order to find DEGs attributable to genetic background. The comparison between treatments was done by comparing the rust control (R), fruit thinning (T), and both (R+T) treatments, with the no control (C) treatment, within cultivars. Contrast tests using the empirical analysis of digital gene expression data (edgeR) as implemented within the CLC Genomics Workbench v9.5.2 were performed using as limiting conditions a Bonferroni correction $p \leq 0.01$, a false discovery rate (FDR) of $p \leq 0.01$, and a fold change (log) cut off of |1| (log2), and only genes significant in both of these tests between cultivars and treatments are referred to as DEGs. The DEGs were used to perform gene ontology (GO) analysis using the Singular Enrichment Analysis (SEA) tool in agriGO v2.0 [188] (http://systemsbiology.cau.edu.cn/agriGOv2/index.php) according to the *Coffea canephora* annotation [134]. Significant GO terms were found using the default FDR $p \leq 0.05$ cutoff value. Venn Diagrams (Bioinformatics & Evolutionary Genomics, Gent, Belgium, http://bioinformatics.psb.ugent.be/webtools/Venn/) were used to compare and visualize the DEGs according to the experimental conditions to represent shared and unique groups.

4.2.4 Candidate genes associated with the phenotypic traits

From the differentially expressed genes involved in the control versus rust control treatments (C vs R) within each cultivar, a Spearman's correlation and Stepwise

regression was performed using the JMP Pro 13.0.0. (SAS Institute Inc. USA) software package. The normalized expression values of the significant DEGs were correlated to each trait described in Table 3 [total leaves (TL), overall condition (OC), rust incidence (RI), rust severity (RS), rust sporulation (RE), and total harvest (TH)].

The correlation analysis was performed, using the pairwise estimation method and a significance *p*-value lower than 0.01. The stepwise analysis was performed using the minimum Bayesian information criterion (BIC) as a stopping rule to select the best adjusted coefficient of determination (closest to 1.0) and FDR of *p*-value lower than 0.05. The DEG was used as the independent variable while the trait was the dependent variable. Any DEG that belonged to a significant GO term, showed significant contribution in the Spearman correlations, and was included in the stepwise regression model, was classified as a candidate gene for that trait.

Linear regressions using 11 to 12 samples from all treatments for each cultivar were used to model the gene expression (normalized counts) as predictors of the percentage of rust sporulation (RE).

4.3 Results

4.3.1 The leaf transcriptome

A total of 5.75 x 10^8 high-quality reads were obtained from the 23 RNA samples after trimming. Approximately 98% of the sequences were between 100-101 bp length after trimming, with 45.0% GC content, and a phred score showing that 98% of the

sequences were higher than 30 (99.9% accuracy in base calling) (Appendix 13). On average, 82% of the sequences aligned to exons and over 93% of the fragments were uniquely mapped to the diploid *C. canephora* [134] reference genome (Figure 22). Overall gene expression with a sequencing depth higher than 0.5X was observed for 4,895 annotated genes (21% of the overall 23,057) (Appendix 14).

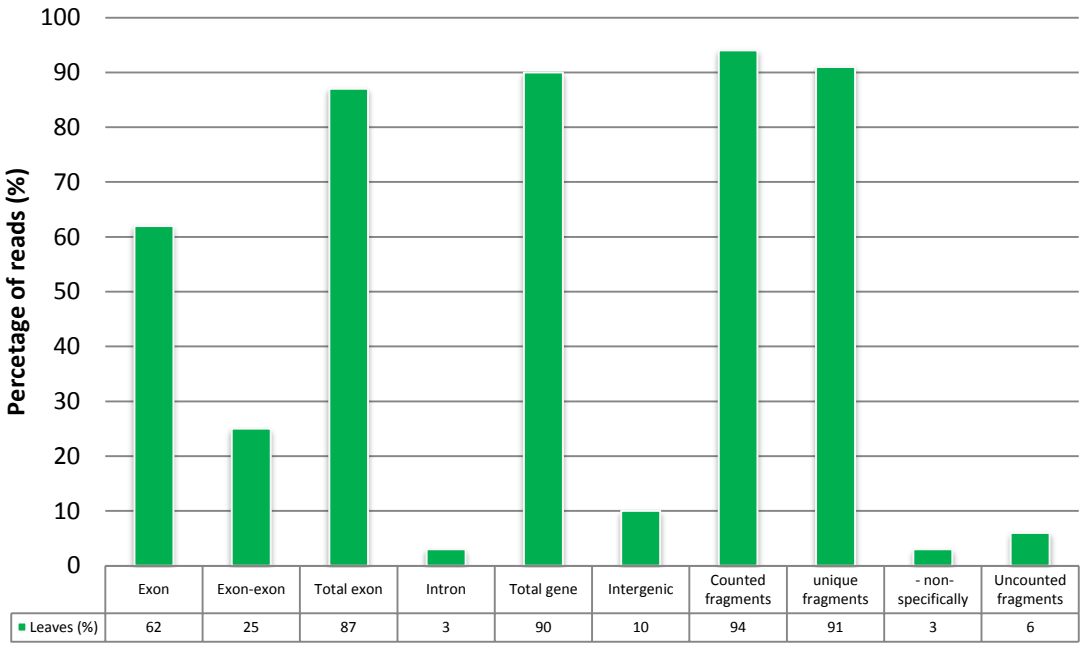


Figure 22. Leaf distribution of the total number of all sequencing reads that passed Illumina's filtering among annotated features across the *Coffea canephora* genome. More details in Appendix 15.

From principal component analysis of the normalized gene expression data (Figure 23) it was determined that the variation between the treatments and cultivars was equal in proportion. Following the interaction of each variable, specific treatment by cultivar comparisons using edgeR were made in order to identify important DEGs within each treatment.

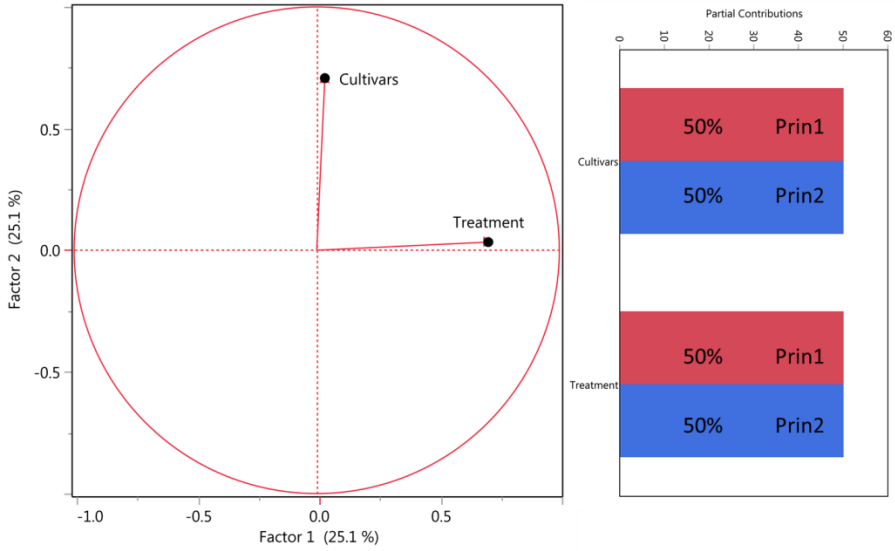


Figure 23. Principal component analysis (PCA) representing the effects of the variables treatment and cultivar.
The first two factors accounted for the 50% of the association between the variables. The partial contribution for the principal components (red and blue colored bars) showed that 50% of the variability was represented by the treatments and 50% by the cultivars, which indicates a dependent interaction between the genes with the genotype and environmental condition.

4.3.2 Comparing cultivars

Comparing both cultivars, 600 DEGs were found to be differentially expressed between the inbred and the hybrid [Appendix 16 shows 136 DEGs with higher sequence depth (>0.5X), without unknown or uncharacterized annotated descriptions]. The inbred had 237 genes that increased in expression when compared against the hybrid, while 363 DEGs increased in the hybrid. The shared and unique DEGs obtained between cultivars, were compared in order to find common DEGs interacting with the treatments (Figure 24). Between cultivars, a higher number of DEGs (either up- or down-regulated) were also found shared within treatments in the hybrid, especially in fruit thinning (T), control

(C), and rust control and fruit thinning (R+T) treatments. A core set of 73/237 DEGs (30.8%) in the inbred and 105/363 DEGs (28.9%) in the hybrid were also found significant across treatments (Figure 24), showing several genes involved in common pathways in a genotype-dependent manner. Since almost all DEGs were found to also be associated to treatment effects, we further analyzed the control (C) versus treatment effects (R, T, or R+T) in edgeR.

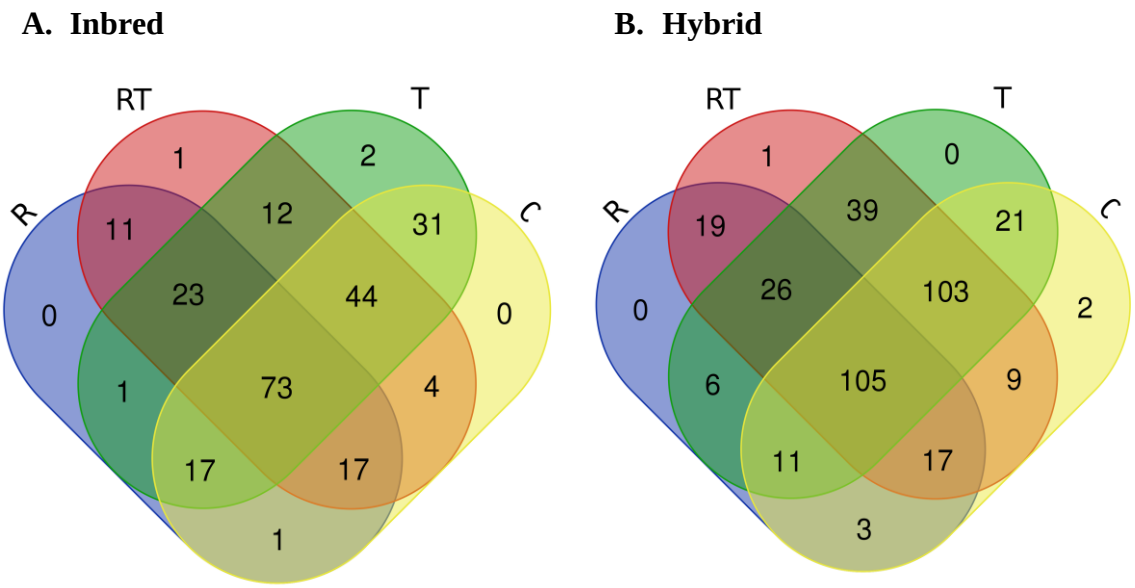


Figure 24. Shared DEGs between treatments in each cultivar.
DEGs are classified according to which treatment or control is increasing the gene expression in the A) inbred and B) hybrid.

4.3.3 Comparing treatments within cultivars

In order to effectively quantify the treatment effects in the edgeR analysis, the normalized values of the genes within each cultivar were used to find differential gene expression. Comparing the control (C) with each treatment (R, T, or R+T), a total of 2,043 unique DEGs were found to be significant within hybrids and inbreds with an FDR and Bonferroni correction of $p \leq 0.01$ (Figure 25). The treatment effect mediated an overall 52% of the gene expression in both cultivars, however, 31% was attributable to overexpression in the inbred whereas 67% was attributable to overexpression in the hybrid. Between 70% to 85% of the DEGs increased in the inbred and hybrid in response to the fruit thinning (T) treatment, respectively. Among the DEGs in the hybrid, the greatest effect was due to the rust control (R) or rust control and fruit thinning (R + T) treatments; in the inbred the highest number of DEGs was in the control (C) treatment.

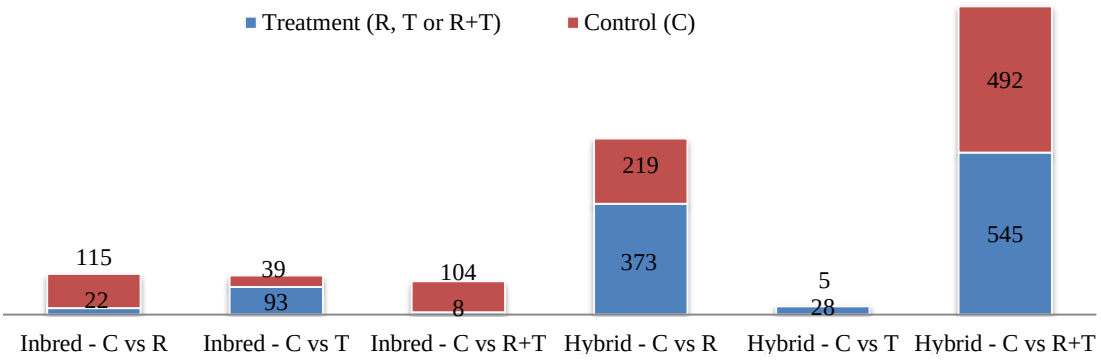


Figure 25. Differentially expressed genes (DEGs) according to the treatment effects in each cultivar.
DEGs are classified according to which treatment (blue color bar) or control (red color bar) resulted in increased gene expression.

4.3.4 Gene ontology analysis

We used all DEGs from each cultivar and each treatment comparison (Figure 24) to find enriched gene ontology (GO) terms (Table 9). When comparing the cultivars, only the inbred showed an enrichment of GO terms (Figure 26). Among the biological functions in the inbred, the terminal GO terms in the pathway with higher importance were chitin catabolic process, cell wall macromolecule catabolic process, regulation of transcription, oxidation-reduction, and multi-organism process. In the molecular function categories (not shown), chitinase activity, heme binding, and electron carrier activity were the most enriched.

The functional GO terms using the treatments and combining the cultivars are summarized in Table 8. The comparison related to the no control vs. rust control (C vs. R) treatments for both cultivars was highly enriched, showing 20 GO terms in the inbred and 30 GO terms in the hybrid. As expected from the previous analyses, the inbred had a higher number of significantly enriched GO terms in the control (C) treatment (Table 9, Comparison #1), while nearly the opposite occurred in the hybrid, which had a higher number of significantly enriched GO terms in the rust control (R) treatment (Table 9, Comparison #8). When examining the treatment of both rust control and fruit thinning compared against the control treatment (C vs R+T) only the hybrid showed an enrichment in GO terms (Table 9, Comparisons #11 and 12).

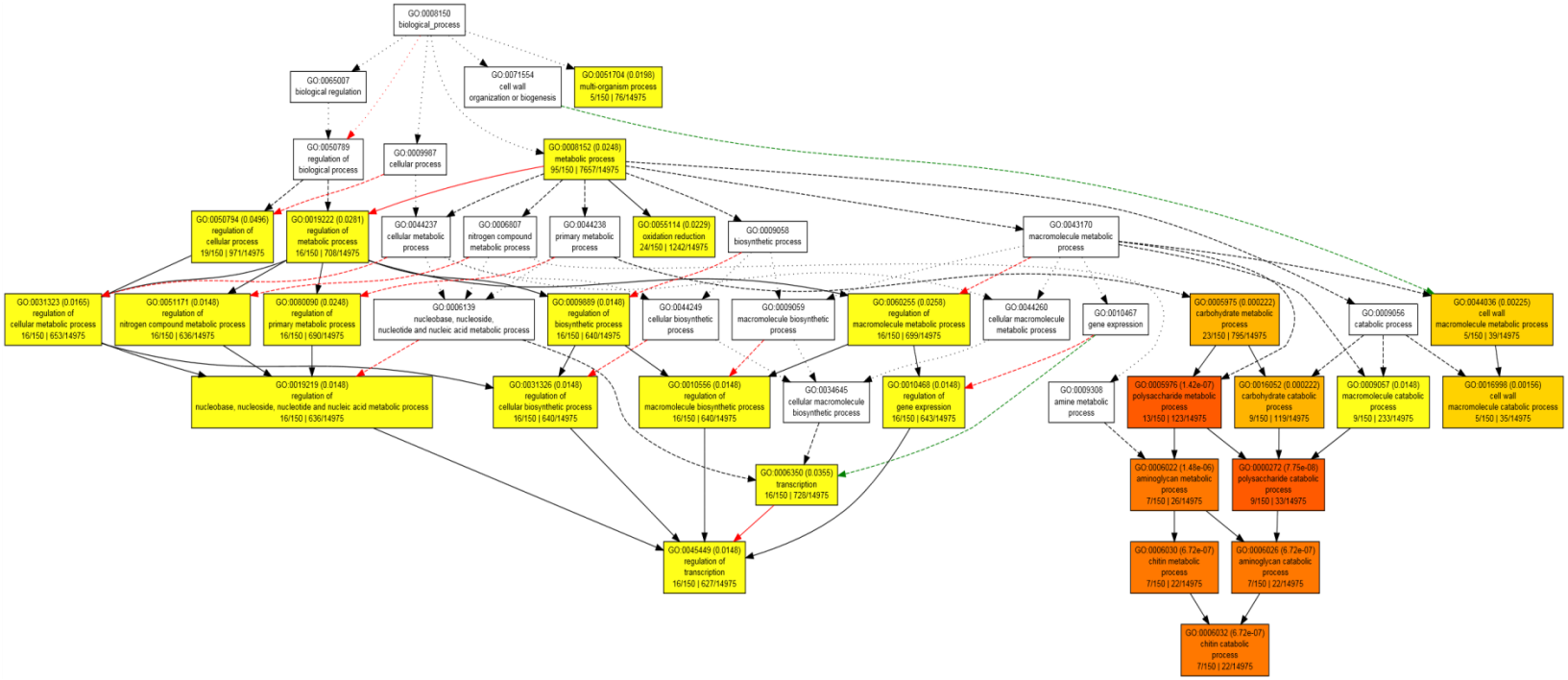


Figure 26. Overall GO terms enriched in biological processes of the inbred when compared to the hybrid.

Squares in dark orange, orange, and yellow determine, higher to lower significantly (FDR p<0.05) enriched terms, respectively.

Table 9. Gene ontology (GO) terms enriched between treatment interactions**.**

For each comparison, the treatments involved, regulation of expression (higher expression in condition…), number of differentially expressed genes (DEGs), and their corresponding gene ontology (GO) final branch terms, are summarized.

The higher regulation indicates what treatment (C, R, T or R+T) resulted in increased expression of the DEGs. The final branch GO term represents the last node in the pathways were almost all other enriched gene ontologies converge. No distinction between the GO term classifications (biological process, molecular function, or cellular component) were used.

#	Cultivar	Comparison	Regulation higher in…	# DEGs	# GO terms	Final branch GO terms	GO description
1	Inbred	C vs R	C	115	18	GO:0016052	carbohydrate catabolic process
						GO:0004497	monooxygenase activity
						GO:0020037	heme binding
2	Inbred	C vs R	R	22	2	GO:0043169	cation binding
3	Inbred	C vs T	C	39	0		
4	Inbred	C vs T	T	93	0		
5	Inbred	C vs R+T	C	104	0		
6	Inbred	C vs R+T	R+T	8	0		
7	Hybrid	C vs R	C	219	5	GO:0055085	transmembrane transport
8	Hybrid	C vs R	R	373	30	GO:0006350	transcription
						GO:0006915	apoptosis
						GO:0006952	defense response
						GO:0005524	ATP binding
						GO:0003777	microtubule motor activity
						GO:0003899	DNA-directed RNA polymerase activity

Table 9 continued…

#	Cultivar	Comparison	Regulation higher in…	# DEGs	# GO terms	Final branch GO terms	GO description
9	Hybrid	C vs T	C	5	0		
10	Hybrid	C vs T	T	28	0		
11	Hybrid	C vs R+T	C	492	35	GO:0006412	translation
						GO:0055114	oxidation reduction
						GO:0006865	amino acid transport
						GO:0009055	electron carrier activity
						GO:0004497	monooxygenase activity
						GO:0020037	heme binding
						GO:0015171	amino acid transmembrane transporter activity
						GO:0005840	ribosome
12	Hybrid	C vs +T	R+T	545	4	GO:0007017	microtubule-based process
						GO:0003777	microtubule motor activity

4.3.5 Candidate genes associated with phenotypic traits

The DEGs that were found significant when comparing the no control versus rust control (C vs R) treatments, were used to find correlations to the phenotypic traits described in Table 3. A total of 906/2,043 annotated genes were found both differentially expressed and correlated to at least one trait. From the annotated genes that were correlated, 496/906 also belonged to a GO term. A total of 144 annotated genes that were chosen by the stepwise regression analysis, were also enriched in a GO term classification (Appendix 17). A total of 785 correlations were found between the DEGs and rust incidence (RI), rust severity (RS), and rust sporulation (RE), but only 88 candidate genes were a) statistically significant (Bonferroni correction and FDR <0.01), b) correlated to RI, RS, and RE, and c) significant in the stepwise regression indistinctly from which trait are predicting (Tables 10 and 11).

From the total number of DEGs associated with the traits, 24 DEGs were found differentially expressed in the inbred cultivar (Table 10, Figure 27). All the candidate genes associated, increased linearly with disease-related parameters (RI, RS, and RE) in the control treatment. The predicted functions of the DEGs found in the inbred were related to oxidation and reduction process, transmembrane transportation, and protein regulation in general.

In the case of the hybrid, all the candidate genes were negatively correlated to phenotypic traits in the rust control treatment, while in the control treatment, there was positive correlation with the disease-related parameters (RI, RS, and RE) (Table 11, Figure 28).

Protein kinases, cations transportation and binding, oxidation and reduction processes, and pathogenesis-related processes, were in general enriched under rust control treatments in the hybrid, which suggests earlier induction of defense responses. Transcription regulation and biosynthesis processing were found to be enriched under no control treatments in the hybrid, showing cellular membrane restructure and cell homeostasis.

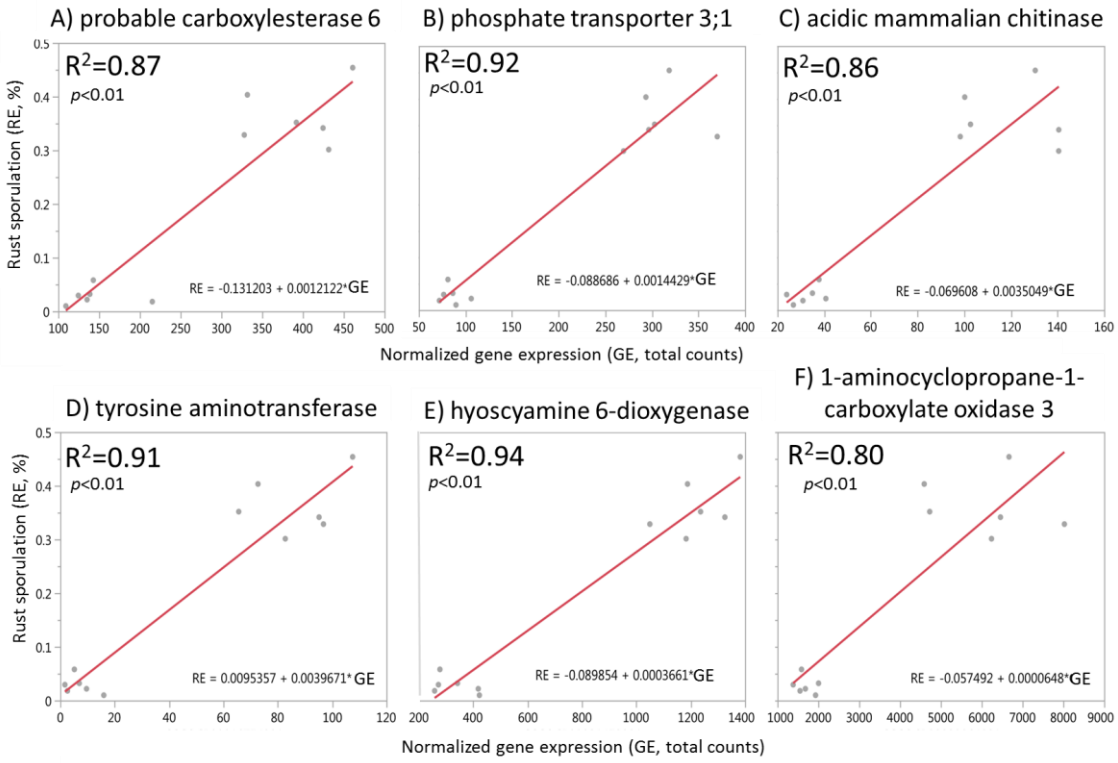


Figure 27. Linear regressions modeling gene expression (GE) as predictors of rust sporulation (RE) in the inbred.
The normalized gene expression (GE, total counts) of the candidate gene is used as predictor of the percentage of rust sporulation (RE). The gene ID corresponds to A) Cc07_g16100, B) Cc04_g09390, C) Cc06_g15430, D) Cc05_g07600, E) Cc05_g10390, and F) Cc05_g02900, as shown in Table 10. Coefficient of determination (R^2), significance level (p<0.01), and prediction equation, are shown in each graph. Data used to plot the linear regression was done using 11 to 12 samples from all treatments.

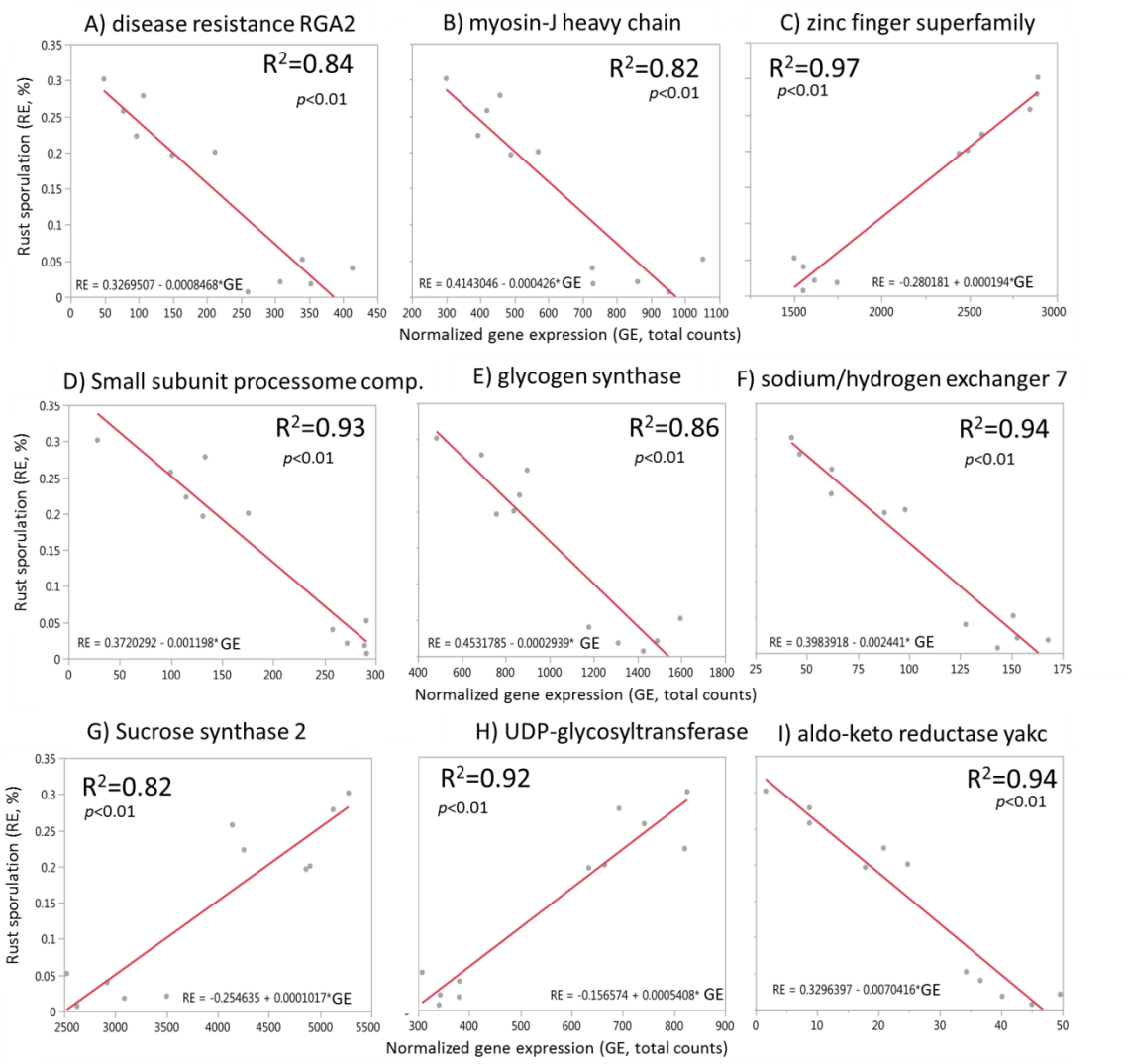


Figure 28. Linear regressions modeling gene expression (GE) as predictors of rust sporulation (RE) in the hybrid.

The normalized gene expression (GE, total counts) of the candidate gene is used as predictor of the percentage of rust sporulation (RE). The gene ID corresponds to A) Cc00_g26280, B) Cc01_g17540, C) Cc04_g15950, D) Cc00_g24200, E) Cc09_g09040, F) Cc11_g05270, G) Cc01_g21050, H) Cc00_g30680, and I) Cc02_g36130, as shown in Table 11. Coefficient of determination (R^2), significance level (p<0.01), and prediction equation, are shown in each graph. Data used to plot the linear regression was done using 11 to 12 samples from all treatments.

Table 10. DEGs associated with disease-related traits in the control treatment in the inbred.

The stepwise regression may have estimate the gene annotated as predicting other traits than the correlations, since the first fits a line quantifying the amount of contribution to the trait value (positive or negative), while the other quantifies the degree to which two variables are related. As observed in the linear regression estimates, the amount of contribution is low (less than |0.01| in 18/24 or 75%), while all the significant correlations ranged between |0.72| and |0.83|. A positive correlation indicates an increase in gene expression as the trait value increases.

Gene annotation (Gene ID[134])	GO term	Spearman correlation			Linear regression	
		Trait	Spearman ρ	Prob>|ρ|	Trait [3]	Estimate
1-aminocyclopropane-1-carboxylate synthase (Cc02_g38530)	catalytic activity	RI	0.75	0.0051	RI	0.0000
phosphate transporter 3;1 (Cc04_g09390)	membrane	RI	0.72	0.0082	RI	0.0003
		RS	0.74	0.0058	RI	0.0003
		RE	0.77	0.0034	RI	0.0003
1-aminocyclopropane-1-carboxylate oxidase 3 (Cc05_g02900)	oxidoreductase activity	RI	0.73	0.0065	RS	0.0000
		RS	0.74	0.0058	RS	0.0000
		RE	0.73	0.0074	RS	0.0000

[3] RI = rust incidence, RS = rust severity, RE = rust sporulation, OC = overall condition, TH = yield, TL = total leaves

Table 10 continued…

Gene annotation (Gene ID[134])	GO term	Spearman correlation			Linear regression	
		Trait	Spearman ρ	Prob>|ρ|	Trait [3]	Estimate
Tyrosine aminotransferase (Cc05_g07600)	-aminocyclopropane--carboxylate synthase activity	RE	0.72	0.0082	RS	0.0004
Putative Hyoscyamine 6-dioxygenase (Cc05_g10390)	oxidoreductase activity	RS	0.72	0.0082	RE	0.0003
		RE	0.81	0.0014	RE	0.0003
Probable anion transporter 3, chloroplastic (Cc06_g01670)	transmembrane transport	RI	0.78	0.0026	OC	-0.0001
		RS	0.80	0.0016	OC	-0.0001
		RE	0.81	0.0014	OC	-0.0001
Putative Stellacyanin (Cc06_g08240)	copper ion binding	RI	0.78	0.0026	OC	-0.0010
		RS	0.76	0.0045	OC	-0.0010
		RE	0.77	0.0034	OC	-0.0010
Putative Acidic mammalian chitinase (Cc06_g15430)	catalytic activity	RE	0.78	0.0026	RE	0.0015
Putative methyltransferase DDB_G0268948 (Cc06_g18640)	methyltransferase activity	RI	0.83	0.0008	TL	0.0167
		RS	0.79	0.0022	TL	0.0167
		RE	0.73	0.0074	TL	0.0167
Putative Probable carboxylesterase 6 (Cc07_g16100)	hydrolase activity	RS	0.74	0.0058	RE	0.0002
		RE	0.83	0.0010	RE	0.0002
Probable calcium-binding protein CML41 (Cc10_g03450)	calcium ion binding	RE	0.73	0.0065	TL	0.0376
serine-type endopeptidase inhibitors (Cc10_g10180)	serine-type endopeptidase inhibitor activity	RE	0.76	0.0045	TH	0.0015

Table 11. DEGs associated with disease-related traits in the rust control or no control treatments in the hybrid.

The stepwise regression may have estimate the gene annotated as predicting other traits than the correlations, since the first fits a line quantifying the amount of contribution to the trait value (positive or negative), while the other quantifies the degree to which two variables are related. As observed in the linear regression estimates, the amount of contribution is low (less than |0.01| in 55/64 or 86% of the DEGs), while all the significant correlations ranged between |0.74| and |0.94|. A negative correlation under up-regulation indicates an increase in gene expression as the trait value increases, while positive correlations under down-regulation indicates an increase in gene expression as the trait value increases.

Regulation	Gene annotation (Gene ID[134])	GO term	Spearman correlation			Linear regression	
			Trait	Spearman ρ	Prob>|ρ|	Trait [4]	Estimate
Up	Putative Small subunit processome component 20 homolog (Cc00_g24200)	binding	RI	-0.76	0.0062	OC	0.0025
Up			RE	-0.87	0.0005	OC	0.0025
Up	Putative Disease resistance protein RGA2 (Cc00_g26280)	defense response	RI	-0.76	0.0062	OC	-0.0007
Up			RE	-0.83	0.0017	OC	-0.0007
Up			RE	-0.81	0.0025	TH	0.1431
Up	Putative Myosin-J heavy chain (Cc01_g17540)	myosin complex	RE	-0.85	0.0008	RE	0.0000

[4] RI = rust incidence, RS = rust severity, RE = rust sporulation, OC = overall condition, TH = yield, TL = total leaves

Table 11 continued…

Regulation	Gene annotation (Gene ID[134])	GO term	Spearman correlation			Linear regression	
			Trait	Spearman ρ	Prob>|ρ|	Trait [4]	Estimate
Up	Oligopeptide transporter 4 (Cc02_g20620)	transmembrane transport	RI	-0.76	0.0062	TL	-0.0160
Up			RE	-0.75	0.0085	TL	-0.0160
Up	Putative Probable 1,4-dihydroxy-2-naphthoate octaprenyltransferase (Cc02_g21250)	integral component of membrane	RI	-0.75	0.0085	RI	0.0000
Up			RE	-0.82	0.0021	RI	0.0000
Up	Putative Probable serine/threonine-protein kinase DDB_G0276461 (Cc02_g27140)	protein serine/threonine kinase activity	RE	-0.76	0.0062	TL	-0.0056
Up	Putative Aldo-keto reductase yakc (Cc02_g36130)	oxidation-reduction process	RI	-0.93	0.0000	RI	-0.0032
Up			RS	-0.87	0.0005	RI	-0.0032
Up			RE	-0.94	0.0000	RI	-0.0032
Up	Putative RING finger and CHY zinc finger domain-containing protein (Cc04_g15950)	zinc ion binding	RE	-0.79	0.0037	TH	-0.0141
Up	Hypothetical protein (Cc04_g16980)	DNA helicase activity	RI	-0.87	0.0005	TL	-0.0022
Up			RS	-0.77	0.0053	TL	-0.0022
Up			RE	-0.93	0.0000	TL	-0.0022
Up	Putative uncharacterized protein (Cc06_g23510)	nucleic acid binding	RE	-0.87	0.0005	RS	-0.0001
Up	ABC transporter C family member (Cc09_g08460)	integral component of membrane	RI	-0.76	0.0062	RS	0.0002
Up			RE	-0.75	0.0085	RS	0.0002
Up	Putative Glycogen synthase (Cc09_g09040)	biosynthetic process	RE	-0.79	0.0037	OC	0.0000
Up	Putative pre-mRNA-processing protein 40A (Cc10_g07620)	protein binding	RI	-0.75	0.0085	RI	0.0004
Up			RE	-0.83	0.0017	RI	0.0004

Table 11 continued…

Regulation	Gene annotation (Gene ID[134])	GO term	Spearman correlation			Linear regression	
			Trait	Spearman ρ	Prob>|ρ|	Trait [4]	Estimate
Up	Armadillo/beta-catenin-like repeat ; C2 calcium/lipid-binding domain (CaLB) protein (Cc10_g12190)	binding	RE	-0.84	0.0013	TL	0.0092
Up	Sodium/hydrogen exchanger 7 (Cc11_g05270)	cation transport	RI	-0.85	0.0010	TL	-0.0028
Up			RS	-0.75	0.0085	TL	-0.0028
Up			RE	-0.91	0.0001	TL	-0.0028
Up	Putative Uridine-cytidine kinase C (Cc11_g08540)	nucleotide binding	RI	-0.80	0.0031	TH	-0.0342
Up			RS	-0.75	0.0073	TH	-0.0342
Up			RE	-0.91	0.0001	TH	-0.0342
Down	Basic endochitinase (Cc00_g14300)	chitinase activity	RE	0.86	0.0006	TH	0.0037
Down	RING/FYVE/PHD zinc finger superfamily protein (Cc00_g23100)	zinc ion binding	RE	0.87	0.0005	RE	-0.0001
Down	UDP-glycosyltransferase 85A3 (Cc00_g30680)	metabolic process	RE	0.86	0.0006	RE	-0.0001
Down	Sucrose synthase 2 (Cc01_g21050)	biosynthetic process	RE	0.80	0.0031	TL	0.0000
Down	Carboxylesterase 1 (Cc02_g03630)	hydrolase activity	RI	0.81	0.0026	RE	-0.0002
Down			RE	0.86	0.0006	RE	-0.0002
Down	Phosphoenolpyruvate carboxylase kinase 2 (Cc02_g08980)	protein serine/threonine kinase activity	RE	0.75	0.0073	RE	0.0000
Down	Flavoprotein WrbA (Cc02_g11960)	FMN binding	RI	0.78	0.0045	RS	0.0000
Down			RE	0.88	0.0003	RS	0.0000

Table 11 continued…

Regulation	Gene annotation (Gene ID[134])	GO term	Spearman correlation			Linear regression	
			Trait	Spearman ρ	Prob>|ρ|	Trait [4]	Estimate
Down	Pathogenesis-related genes transcriptional activator PTI5 (Cc02_g14240)	transcription factor activity, sequence-specific DNA binding	RI	0.80	0.0031	RE	0.0001
Down			RE	0.81	0.0026	RE	0.0001
Down	Arginine decarboxylase (Cc02_g16860)	arginine decarboxylase activity	RE	0.74	0.0098	RS	0.0000
Down	phosphate transporter 3;1 (Cc04_g09390)	membrane	RI	0.81	0.0026	RI	0.0003
Down			RS	0.74	0.0098	RI	0.0003
Down			RE	0.80	0.0031	RI	0.0003
Down	Putative Stellacyanin (Cc06_g08240)	copper ion binding	RI	0.75	0.0085	OC	-0.0010
Down			RS	0.74	0.0098	OC	-0.0010
Down			RE	0.81	0.0026	OC	-0.0010
Down	PeroxiRedoxin-2B (Cc06_g09990)	cell Redox homeostasis	RE	0.77	0.0053	RI	0.0000
Down	Nuclear transport factor 2 (Cc07_g00920)	transport	RI	0.75	0.0085	RE	0.0004
Down			RE	0.91	0.0001	RE	0.0004
Down	Dihydrodipicolinate synthase, chloroplastic (Cc08_g04960)	lyase activity	RI	0.82	0.0021	RI	0.0074
Down			RE	0.87	0.0005	RI	0.0074
Down	V-type proton ATPase 16 kDa proteolipid subunit (Cc10_g02030)	ATP synthesis coupled proton transport	RE	0.80	0.0031	TL	0.0018
Down	Probable calcium-binding protein CML41 (Cc10_g03450)	calcium ion binding	RE	0.82	0.0021	TL	0.0376

Table 11 continued…

Regulation	Gene annotation (Gene ID[134])	GO term	Spearman correlation			Linear regression	
			Trait	Spearman ρ	Prob>|ρ|	Trait [4]	Estimate
Down	Probable trehalose-phosphate phosphatase E (Cc10_g07410)	catalytic activity	RI	0.87	0.0005	RI	0.0000
Down			RS	0.84	0.0013	RI	0.0000
Down			RE	0.94	0.0000	RI	0.0000
Down	Putative LRR receptor-like serine/threonine-protein kinase FLS2 (Cc11_g06630)	protein binding	RE	0.79	0.0037	RI	0.0010
Down			RE	0.78	0.0045	TH	0.0114
Down	40S ribosomal protein S27-1 (Cc11_g14900)	structural constituent of ribosome	RI	0.88	0.0003	TL	-0.0014
Down			RS	0.82	0.0021	TL	-0.0014
Down			RE	0.85	0.0010	TL	-0.0014

4.4 Discussion

4.4.1 Leaf transcriptome

The present study provides an overview of the transcriptome of the coffee leaf response induced by two types of management to control coffee leaf rust disease (CLR). Our annotation showed that 89% of the *C. arabica* sequences were aligned to the *C. canephora* genome which was higher than a *de novo* assembly [79]. Since the gene expression (GE) annotation was selected in our study, we limit our research to the known coding genes in the *C. canephora*, missing those from the *C. eugenoides*.

According to our data, the total variation of the normalized gene expression was equally attributed to the cultivars and treatments, which represent genotype by environment interaction. Since both cultivars and four different treatments were involved in changes in gene expression, controlling one parameter (i. e. cultivars), led us to a better statistical approach to dissect and quantify differential gene expression.

4.4.2 Treatment and cultivar DEGs

The number of DEGs obtained when comparing cultivars, showed a higher content of DEGs involved in treatment interactions within the hybrid. When comparing the treatments effects (intended to reduce disease-related stresses) against the no-treatment negative control, the hybrid showed a greater number of DEGs when less stress occurred. The presence of a higher number of DEGs under less stress, suggests

higher tolerance or plasticity, as has been shown to be of the case in interspecific and intraspecific *Coffea* hybrids [149]. Another reasonable explanation of higher abundance of differentially expressed genes under rust control treatments in the hybrid, is that the fungicide acted as an elicitor of biotic stress, as shown by Monteiro, et al. [227] using phosphite products.

Contrary to the hybrid, when analyzing the inbred more DEGs were found in the treatment with more stress (i.e. the control treatment), suggesting the accumulation of homozygous allele combinations fixed by selfing and inbreeding, acting in a conserved pathway as response of stress [228].

4.4.3 Treatment and cultivar GO

We used the DEGs identified in the treatments to find enriched gene ontology (GO) terms. By using the t DEGs from the various treatments to find GO terms, we were able to verify that the hybrid and inbred had different responses to the different treatments, which explains the differences in gene expression observed when comparing cultivars. The up-regulated DEGs the hybrid cultivar did not belong to any enriched GO terms. The probable reason of not finding enriched GO terms for the up-regulated genes in the hybrid, suggests a higher distribution of DEGs coming from unrelated pathways and thus no specific enrichment GO terms were found. The up-regulated DEGs in the control treatment in the inbred were related to an increase of disease-related stress, however, the enriched GO terms were related to carbohydrate, monoxygenase, and heme binding processes, which are not directly associated with a defense reaction. The

enriched GO terms in the hybrid rust control treatment were related to defense response and apoptosis, which are associated with host-pathogen interactions [52].

4.4.4 Candidate DEGs associated with traits

The most related-to-CLR-disease parameters that were altered in the field were the rust incidence (RI), rust severity (RS), and rust sporulation (RE) [180]. The treatments that compared the no control with the rust control (C vs R) treatments, showed a higher number of significant enriched GO terms in both cultivars. Considering that the spray application reduced an average 12%, 3%, and 27% of rust incidence, rust severity, and rust sporulation in both cultivars (Table 3), respectively, spray application was also useful to uncover variation in gene expression. The correlation of the phenotype with the transcriptome profiles revealed that the management of CLR disease enriches certain GO terms. From 100 candidate DEGs associated with traits, 88 were correlated with rust incidence, rust severity, and rust sporulation in a genotype-dependent manner.

The overall transcriptome information and candidate genes analysis revealed two different types of defense response. A) In the case of the inbred, the defense response was highly oriented into carbohydrate metabolism and oxidative burst signaling, associated with a hypersensitive response (HR). However, B) in the case of the hybrid, higher number of defense-related and recovery proteins are reported to be more related to a systemic acquired resistance (SAR). Even though both cultivars are susceptible to CLR, the hybrid showed a 4%, 1%, and 5%, overall reduction in rust incidence, rust

severity, and rust sporulation, respectively, when compared to the inbred (Table 3), which suggests some tolerance level and transcriptomic plasticity.

The inbred suggests an early hypersensitive response (HR) induced by the oxidative burst. Expression of genes encoding a putative hyoscyamine 6-dioxygenase, a putative stellacyanin, a tyrosine aminotransferase, a putative acidic mammalian chitinase, a putative methyltransferase DDB_G0268948, and other proteins related to oxidoreductase activities increased with RE (Table 10, Figure 27). The hyoscyamine 6-dioxygenase (Table 10, Figure 27-E) has been found to be expressed early in HR responses of potato under *Phytophthora infestans* attack [229]. The stellacyanins have stronger oxidation potential than other cupredoxins [230], which represent more sensitive signaling under ROS accumulation. Tyrosine aminotransferase (Figure 27-D) is involved in tocopherol synthesis [231] and other benzylisoquinoline alkaloids [232] which are natural antioxidants. Chitinase activity, the presence of a putative anion transporter, and calcium influx, are signatures that an early oxidative burst event associated with the HR was ongoing, as has also been suggested in the citrus CitEST database that considered a broad combinations of limiting factors, phenological stages, and tissues [233]. The putative methyltransferase DDB_G0268948, has been also found highly expressed under Redox activity after *Pseudomonas syringae* pv *tomato* inoculation in a resistant tomato [234] and in *Paulownia* stress tolerance [235].

The inbred had candidate DEGs related to carbohydrate catabolic process and monooxygenase activity (Table 9, #1). Expression of genes encoding proteins such as 1-aminocyclopropane-1-carboxylate (ACC) synthase and oxidase (Table 10, Figure 27-F)

increased with rust incidence (RI). These proteins are the precursors for ethylene production [236]. Since ROS (specially ozone) activates ethylene production to induce ethylene-mediated cell death, a hypersensitive response (HR) is suggested to happen in the inbred-CLR interaction mediated by pathogenesis-related proteins [237].

By the other hand in the hybrid, one of the candidate DEGs found differentially expressed was the disease resistance gene RGA2 (Table 11, Figure 28-A). The same type of gene was reported by Florez, et al. [201] in coffee-CLR interaction, which also is related to CLR recognition by the putative LRR receptor-like serine/threonine-protein kinase FLS2 (Table 11). The function of RGA2 is related to signal perception and transduction under a systemic acquired resistance (SAR) interaction. From our findings, the expression of RGA2 decreased with less rust incidence (RI) and rust sporulation (RE) in the hybrid. The same pattern was exhibited by an ABC transporter C family member 1 gene (Table 11), which is involved in synthesis and transport of antimicrobial metabolites [238]. The arginine decarboxylase (Table 11) involved in signal perception and transduction was down-regulated in the hybrid as well, showing active gene regulation related to host-pathogen interactions.

The transmembrane transport and protein structure GO categories were highly represented during biotic stress (Table 9, #7 and 11). The higher expression of the oligopeptide transporter 4 correlated with rust incidence (RI) and rust sporulation (RE) in the hybrid (Table 11) indicates plant and fungus interactions, as also reported in grape [239]. Genes containing RING finger domains and chitinases (Table 11, Figure 28-C), also found by Guzzo, et al. [238], are in charge of regulating protein degradation and

antimicrobial proteins, respectively. A sodium/hydrogen exchanger (Figure 28-F) and armadillo/beta-catenin-like repeat C2 calcium/lipid-binding domain (CaLB) protein (Table 11), have been shown to be involved in submergence tolerance in rice [240]. According to Li, et al. [241], the armadillo/beta-catenin-like repeat domains are also involved in transcriptional regulation, protein degradation, chromatin remodeling, and cytoskeletal regulation under stress, suggesting that intense regulation and remodeling of the cell wall was occurring under CLR infestation, and may explain the lower rust incidence (RI) and rust sporulation (RE) (Table 3).

Growth and developmental proteins such as UDP-glycotransferases, putative uridine/cytidine kinase, and a putative myosin-J heavy chain protein (Table 11, Figure 28), were also correlated with rust sporulation (RE) in the hybrid. Increases in UDP-glycotransferases have been associated with the addition of sugars to plant hormones under ROS cascades, modulating indole-3-butiric acid (IBA) homoeostasis and inducing water stress tolerance [242]. The uridine/cytidine kinase (EC 2.7.1.48) is in charge of nucleoside degradation and salvage, which supports active growth [243]. Reduction of its expression under biotic stress, suggests a defense mechanism to regulate purine and pyrimidine metabolism in order to reduce cell proliferation. The myosins proteins are in charge of the reorganization and polarization of actin filaments inside the cell [244]. Since at the penetration sites, the first barrier to limit the pathogens growth and infestation is by reordering the cytoskeleton and organelles, an increase in the expression of myosin under stress suggests a cytological defense mechanism to limit CLR penetration and expansion.

The protein biosynthesis and maturation process were also affected under higher CLR disease (Table 9, #11). Expression of genes encoding a putative small subunit processome component 20 homolog (Figure 28-D) and a putative pre-mRNA-processing protein 40A (Table 11) decreased with increasing rust incidence (RI) and rust sporulation (RE) in the hybrid. The small ribosomal subunit (SSU) processome is a nuclear large ribonucleoprotein (RNP) required for processing the precursors of the 18S small subunit RNA of the ribosome, in charge of rRNA transcription and ribosome assembly [245]. The putative pre-mRNA-processing protein 40A interacts with a mediator complex 35 as co-regulators of protein transcription [246]. If the reduction in gene expression for these two proteins resulted in decreased protein activity, this suggests that *de novo* biosynthesis of proteins was regulated in the nuclei in response to CLR stress, perhaps to stop cell growth and facilitate the metabolism change in order to mitigate the stress and later recovery [247].

Expression of genes related to anti-oxidation processes were also increased with CLR stress in the hybrid (Table 9, #11). For example, expression of a peroxiRedoxin-2B was found to be correlated with RE (Table 11). The peroxiRedoxin-2B, was also reported by Margaria, et al. [248] in grape under phytoplasma attack, which was one of the first chloroplastic enzymes involved in the response to oxidative stress and recovery to steady-state. Expression of the genes coding for the aldo-keto reductase yakc proteins (Table 11, Figure 28-I) increased with RI, RS, and RE, and this has also been seen in response to drought stress in maize [249]. The aldo-keto reductase yakc proteins are activated in an abscisic acid (ABA)-dependent way, revealing that hormone signaling

and oxidation-reduction activity were ongoing during CLR attack in the hybrid in a manner consistent with systemic acquired resistance (SAR).

The overall transcriptome and correlation analyses, suggests that the hybrid had a broader response to the coffee leaf rust (CLR) disease. The quantitative expression of genes related to stress under higher rust sporulation in the treatment without any management, but also apoptosis and a qualitative response under less stressed treatments, evidence that the genotype has the ability to regulate the gene transcription, but also plenty of genes that network. The higher tolerance of the hybrid suggests being part of the vigor, not only expressed in less CLR disease, but also higher productivity under good management practices.

Further research has to be addressed in order to first validate the set of candidate genes obtained in this project, and later expanding to other genotypes, managements, and environments.

CHAPTER V

CONCLUSIONS AND FURTHER RESEARCH

Coffee is and will continue to be, one of the most important commodity crops in the world. Because of the beverage qualities for health, stimulation, and pleasure, coffee is part of the human diet. Nevertheless, the coffee marketing chain is restricted to availability, growth, and properties of the bean, which relies on the producing countries and uncontrollable environmental conditions. The challenge to supply enough coffee to growing world demand requires comprehending the interactions of fruit yield, bean chemical composition, and environmental conditions. Since the only controllable aspect at the farm level are the type of plant [genotype (G)] and management agronomic practices [management (M)], the transcriptome study was shown to be a reasonable approach to find relevant genes that can be participating in the response.

In Chapter II we found that small G x M interactions occurred under CLR stress. After the monthly effect in the disease parameters, rust control had the major influence on experiment wise variation (20%). The spray application reduced rust incidence (RI) 12% overall and rust sporulation (RE) was reduced by 27%. These are useful determinants in the continuation of the CLR cycle. Related to the beverage quality, the rust control treatment alone reduced the clean cup and overall scores, while it increased musty dusty aroma in small proportions; however, jointly with fruit thinning, it reduced the stale aroma and acetaminophen in a genotype dependent manner. Using both

cultivars under rust control and fruit thinning treatments revealed that G x M interactions under biotic stress can be sensitively detected in the bean chemistry and sensory profiles.

Following our first observations, in Chapter III, the overall transcriptome of the mature and immature fruits also showed sensitive gene expression differences. The functional annotation of a total 471 differentially expressed genes (DEGs) related to the cultivars, maturity stages, and combined effects of rust control and thinning treatments into 19 GO terms, reflects that the coffee bean's gene expression can be modified by management practices. Our analysis used a conservative approach to find statistical differences and narrow the set of genes; however, many other annotated genes and biological functions can be expected as part of the transcriptome network complexity. The quantity of DEGs also reflects possible candidate genes to be validated in the future as function of their metabolic pathways, especially in carbohydrate and fatty acid biosynthesis and metabolism. Since several GO terms and DEGs were found correlated to quality precursors, it is expected that the quality of a cup of coffee can be also explained as a function of the gene expression under certain cultivation conditions.

The cultivar effect was further explored in Chapter IV using the leaf transcriptome. Since the hybrid previously showed more DEGs in enriched GO terms related to carbohydrate and lipid metabolism in the beans than the inbred, it suggests that a greater metabolic plasticity can serve as a clue of higher vigor and CLR tolerance. Several pathways and genes (characterized or not) related to defense responses in the leaves, seem to be in accordance with a quantitative or systemic acquired resistance (SAR) in the hybrid and a qualitative or hypersensitive response (HR) in the inbred.

Using biological and statistical analysis, it was possible to find 88 candidate genes with significant correlation and effect with rust incidence (RI), rust severity (RS), and rust sporulation (RE).

The interconnection between tissues is in accordance to a systemic response over all the plant. The biotic stress and health condition of the plant interacted in the leaf tissue by affecting the ability of the plant to tolerate the disease. In the fruits, the stress response perceived was translated into a defense mechanism in the cell walls to transform fatty acids and carbohydrates stored, which effect was the change in volatiles found in the green beans and after roasting. From the plant to the cup, the history of how the coffee bean was managed was recorded.

Considering the overall treatment effects in this research, we consider that studying genotype by management interactions clearly augments the comprehension of genome plasticity and possible signal networks related to stress response and quality. A better understanding of gene networks and dynamics can be translated into better accuracy and control of complex quantitative traits during selection processes in a breeding program. The transcriptome information can be used therefore to explain several pathways related to biotic stress response and quality, in order to breed for more durable resistance without decreasing beverage quality.

REFERENCES

1. ITC, TRADEMAP: Trade statistics for international business development. **2016**.

2. DaMatta, F. M.; Ramalho, J. D. C., Impacts of drought and temperature stress on coffee physiology and production: a review. *Brazilian Journal of Plant Physiology* **2006,** *18*, 55-81.

3. Redden, R., New approaches for crop genetic adaptation to the abiotic stresses predicted with climate change. *Agronomy* **2013,** *3* (2), 419-432.

4. Kilambo, D. L.; Reuben, S. O.; Mamiro, D., Races of *Hemileia vastatrix* and variation in pathogenicity of *Colletotrichum kahawae* isolates to compact coffee genotypes in Tanzania. *J. Plant Stud.* **2013,** *2* (2), 95.

5. Leroy, T.; De Bellis, F.; Legnate, H.; Kananura, E.; Gonzales, G.; Pereira, L. F.; Andrade, A. C.; Charmetant, P.; Montagnon, C.; Cubry, P., Improving the quality of African robustas: QTLs for yield-and quality-related traits in *Coffea canephora*. *Tree Genetics & Genomes* **2011,** *7* (4), 781-798.

6. van der Vossen, H.; Bertrand, B.; Charrier, A., Next generation variety development for sustainable production of arabica coffee (*Coffea arabica* L.): a review. *Euphytica* **2015,** *204* (2), 243-256.

7. Wikström, N.; Kainulainen, K.; Razafimandimbison, S. G.; Smedmark, J. E.; Bremer, B., A revised time tree of the asterids: establishing a temporal framework for evolutionary studies of the coffee family (Rubiaceae). *PloS one* **2015,** *10* (5), e0126690.

8. Guyot, R.; Lefebvre-Pautigny, F.; Tranchant-Dubreuil, C.; Rigoreau, M.; Hamon, P.; Leroy, T.; Hamon, S.; Poncet, V.; Crouzillat, D.; De Kochko, A., Ancestral synteny

shared between distantly-related plant species from the asterid (*Coffea canephora* and *Solanum* sp.) and rosid (*Vitis vinifera*) clades. *BMC Genomics* **2012,** *13* (1), 1.

9. Davis, A. P.; Tosh, J.; Ruch, N.; Fay, M. F., Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of *Coffea*. *Bot. J. Linn. Soc.* **2011,** *167* (4), 357-377.

10. Chevalier, A. *Les caféiers du globe. Fasc. 3: Systématique des caféiers et faux-caféiers maladies et insectes nuisibles*; Paul Lechevalier: 1947.

11. Dentan, E.; Clifford, M.; Willson, K., Coffee: botany, biochemistry and production of beans and beverage. *Coffee: botany, biochemistry and production of beans and beverage* **1985**.

12. Rodrigues, C. J.; Bettencourt, A. J.; Rijo, L., Races of the pathogen and resistance to coffee rust. *Annu. Rev. Phytopathol.* **1975,** *13*.

13. Davis, A. P.; Govaerts, R.; Bridson, D. M.; Stoffelen, P., An annotated taxonomic conspectus of the genus Coffea (Rubiaceae). *Bot. J. Linn. Soc.* **2006,** *152* (4), 465-512.

14. Hamon, P.; Grover, C. E.; Davis, A. P.; Rakotomalala, J.-J.; Raharimalala, N. E.; Albert, V. A.; Sreenath, H. L.; Stoffelen, P.; Mitchell, S. E.; Couturon, E.; Hamon, S.; de Kochko, A.; Crouzillat, D.; Rigoreau, M.; Sumirat, U.; Akaffou, S.; Guyot, R., Genotyping-by-sequencing provides the first well-resolved phylogeny for coffee (*Coffea*) and insights into the evolution of caffeine content in its species: GBS coffee

phylogeny and the evolution of caffeine content. *Molecular Phylogenetics and Evolution* **2017,** *109*, 351-361.

15. Clifford, M. N., *Coffee: botany, biochemistry and production of beans and beverage*. Springer Science & Business Media: 2012.

16. Cramer, P. J. S.; Wellman, F. L., *A review of literature of coffee research in Indonesia*. SIC Editorial, Inter-American Institute of Agricultural Sciences Turrialba, Costa Rica: 1957.

17. Ukers, W. H., *All about coffee*. Tea and Coffee Trade Journal Company: 1922.

18. Bettencourt, A.; Noronha-Wagner, M., Genetic factors conditioning resistance of *Coffea arabica* L. to *Hemileia vastatrix* Berk. et Br. *Agron. Lusit.* **1971,** *31*, 285-292.

19. Yuyama, P. M.; Júnior, O. R.; Ivamoto, S. T.; Domingues, D. S.; Carazzolle, M. F.; Pereira, G. A. G.; Charmetant, P.; Leroy, T.; Pereira, L. F. P., Transcriptome analysis in *Coffea eugenioides*, an Arabica coffee ancestor, reveals differentially expressed genes in leaves and fruits. *Molecular Genetics and Genomics* **2016,** *291* (1), 323-336.

20. Cenci, A.; Combes, M.-C.; Lashermes, P., Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. *Plant molecular biology* **2012,** *78* (1-2), 135-145.

21. Combes, M. C.; Dereeper, A.; Severac, D.; Bertrand, B.; Lashermes, P., Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasted temperatures. *New phytologist* **2013,** *200* (1), 251-260.

22. Klein, P. E.; Murray, S. C.; Solano, W.; Montagnon, C.; Schilling, T.; Bertrand, B., An extensive study of the genetic diversity of Arabica coffee reveals two subpopulations in Ethiopia and a cultivated subpopulation in Yemen In *ASIC*, Yunnan, China, 2016.

23. Maluf, M. P.; Silvestrini, M.; Ruggiero, L. M. d. C.; Guerreiro Filho, O.; Colombo, C. A., Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. *Scientia Agricola* **2005,** *62* (4), 366-373.

24. Ghini, R.; Bettiol, W.; Hamada, E., Diseases in tropical and plantation crops as affected by climate changes: current knowledge and perspectives. *Plant pathology* **2011,** *60* (1), 122-132.

25. Lin, B. B., Resilience in agriculture through crop diversification: adaptive management for environmental change. *BioScience* **2011,** *61* (3), 183-193.

26. Barquero-Miranda, M., Las variaciones climáticas en el incremento inusual de la roya del cafeto. *Revista Informativa* June, 2013, pp 1-7.

27. Bettencourt, A. J.; Rodrigues, C. J., Principles and practice of coffee breeding for resistance to rust and other diseases. In *Coffee Agronomy*, Clarke, R. J.; Macrae, R., Eds. Elsevier Applied Science Publishers: London and New York, 1988.

28. Talhinhas, P.; Batista, D.; Diniz, I.; Vieira, A.; Silva, D. N.; Loureiro, A.; Tavares, S.; Pereira, A. P.; Azinheira, H. G.; Guerra-Guimarães, L. The coffee leaf rust pathogen *Hemileia vastatrix*: one and a half centuries around the tropics *Mol. Plant Pathol.* [Online], 2016.

29. Zamora, Q.; Cooper, J.; Lipper, L. M.; Zilberman, D.; Mardoqueo González, J.; Montaldo, A.; Echeverri, J.; Zamora, L.; Osuna Canizalez, F. d. J.; Montilla de Bravo, I. *Manual de recomendaciones para el cultivo del café*; 9977550212; Instituto del Café de Costa Rica.: 1998.

30. Tavares, S.; Ramos, A. P.; Pires, A. S.; Azinheira, H. G.; Caldeirinha, P.; Link, T.; Abranches, R.; Silva, M. d. C.; Voegele, R. T.; Loureiro, J., Genome size analyses of Pucciniales reveal the largest fungal genomes. *Frontiers in plant science* **2014,** *5*, 422.

31. D'hondt, L.; Höfte, M.; Van Bockstaele, E.; Leus, L., Applications of flow cytometry in plant pathology for genome size determination, detection and physiological status. *Mol. Plant Pathol.* **2011,** *12* (8), 815-828.

32. McTaggart, A. R.; Shivas, R. G.; Nest, M. A.; Roux, J.; Wingfield, B. D.; Wingfield, M. J., Host jumps shaped the diversity of extant rust fungi (Pucciniales). *New Phytologist* **2016,** *209* (3), 1149-1158.

33. Fernandes, R. d. C.; Evans, H. C.; Barreto, R. W., Confirmation of the occurrence of teliospores of *Hemileia vastatrix* in Brazil with observations on their mode of germination. *Fitopatol. Bras.* **2009,** *34* (2), 108-113.

34. Carvalho, C. R.; Fernandes, R. C.; Carvalho, G. M. A.; Barreto, R. W.; Evans, H. C., Cryptosexuality and the genetic diversity paradox in coffee rust, *Hemileia vastatrix*. *PLoS One* **2011,** *6* (11), e26387.

35. Coutinho, T.; Rijkenberg, F.; Van Asch, M., Teliospores of *Hemileia vastatrix*. *Mycological Research* **1995,** *99* (8), 932-934.

36. Chinnappa, C.; Sreenivasan, M., Cytology of *Hemileia vastatrix*. *Caryologia* **1968,** *21* (1), 75-82.

37. Silva, M. C.; Várzea, V.; Guerra-Guimarães, L.; Azinheira, H. G.; Fernandez, D.; Petitot, A. S.; Bertrand, B.; Lashermes, P.; Nicole, M., Coffee resistance to the main diseases: leaf rust and coffee berry disease. *Braz J Plant Physiol* **2006,** *18*.

38. Fernandez, D.; Tisserant, E.; Talhinhas, P.; Azinheira, H.; Vieira, A.; PETITOT, A. S.; Loureiro, A.; Poulain, J.; Da Silva, C.; SILVA, M., 454-pyrosequencing of Coffea arabica leaves infected by the rust fungus Hemileia vastatrix reveals in planta-expressed pathogen-secreted proteins and plant functions in a late compatible plant–rust interaction. *Molecular plant pathology* **2012,** *13* (1), 17-37.

39. Junior, R.; Bettencourt, C.; AJ Rijo, L., Races of the pathogen and resistance to coffee rust. *Annual Review of Phytopathology (EUA) v. 13 p. 49-70* **1975**.

40. D'Oliveira, B., Selection of coffee types resistant to the *Hemileia* leaf rust. *Coffee and Tea Industry Flavor Field* **1958,** *8*, 112-20.

41. Flor, F., Inheritance of pathogenicity in *Melampsora lini*. *Phytopathology* **1942,** *32*, 653-669.

42. Parlevliet, J. E.; Zadoks, J. C., The integrated concept of disease resistance: A new view including horizontal and vertical resistance in plants. *Euphytica* **1977,** *26* (1), 5-21.

43. Sera, G. H.; Sera, T.; Ito, D. S.; Azevedo, J. A. d.; Mata, J. S. d.; Dói, D. S.; Ribeiro Filho, C.; Kanayama, F. S., Resistance to leaf rust in coffee carrying *SH3* gene and others *SH* genes. *Brazilian Archives of Biology and Technology* **2007,** *50*, 753-757.

44. Alkimim, E. R.; Caixeta, E. T.; Sousa, T. V.; Pereira, A. A.; de Oliveira, A. C. B.; Zambolim, L.; Sakiyama, N. S., Marker-assisted selection provides arabica coffee with genes from other Coffea species targeting on multiple resistance to rust and coffee berry disease. *Molecular Breeding* **2017,** *37* (1), 6.

45. Romero, G.; Vásquez, L. M.; Lashermes, P.; Herrera, J. C., Identification of a major QTL for adult plant resistance to coffee leaf rust (*Hemileia vastatrix*) in the natural Timor hybrid (*Coffea arabica* x *C. canephora*). *Plant breeding* **2014,** *133* (1), 121-129.

46. Romero, G.; Alvarado, G.; Ligarreto, G.; Galeano, N. F., Partial resistance to leaf rust (*Hemileia vastatrix*) in coffee (*Coffea arabica* L.): genetic analysis and molecular characterization of putative candidate genes. *Molecular breeding* **2010,** *25* (4), 685-697.

47. Jones, J. D.; Dangl, J. L., The plant immune system. *Nature* **2006,** *444* (7117), 323-329.

48. Dong, X.; Jiang, Z.; Peng, Y.-L.; Zhang, Z., Revealing shared and distinct gene network organization in *Arabidopsis* immune responses by integrative analysis. *Plant Physiol.* **2015,** *167* (3), 1186-1203.

49. Nurnberger, T.; Lipka, V., Non-host resistance in plants: new insights into an old phenomenon. *Mol. Plant Pathol.* **2005,** *6* (3), 335-345.

50. Kump, K. L.; Bradbury, P. J.; Wisser, R. J.; Buckler, E. S.; Belcher, A. R.; Oropeza-Rosas, M. A.; Zwonitzer, J. C.; Kresovich, S.; McMullen, M. D.; Ware, D.; Balint-Kurti, P. J.; Holland, J. B., Genome-wide association study of quantitative

resistance to southern leaf blight in the maize nested association mapping population. *Nature Genetics* **2011,** *43*, 163.

51. Bazakos, C.; Hanemian, M.; Trontin, C.; Jiménez-Gómez, J. M.; Loudet, O., New strategies and tools in quantitative genetics: how to go from the phenotype to the genotype. *Annu. Rev. Plant Biol.* **2017,** *68* (1), 435-455.

52. Kushalappa, A. C.; Yogendra, K. N.; Karre, S., Plant innate immune response: qualitative and quantitative resistance. *Critical Reviews in Plant Sciences* **2016,** *35* (1), 38-55.

53. Kou, Y.; Wang, S., Broad-spectrum and durability: understanding of quantitative disease resistance. *Current Opinion in Plant Biology* **2010,** *13* (2), 181-185.

54. Miller, R. N. G.; Costa Alves, G. S.; Van Sluys, M.-A., Plant immunity: unravelling the complexity of plant responses to biotic stresses. *Ann. Bot.* **2017,** *119* (5), 681-687.

55. Kemen, A. C.; Agler, M. T.; Kemen, E., Host–microbe and microbe–microbe interactions in the evolution of obligate plant parasitism. *New Phytologist* **2015,** *206* (4), 1207-1228.

56. Holoch, D.; Moazed, D., RNA-mediated epigenetic regulation of gene expression. *Nature Reviews Genetics* **2015,** *16* (2), 71-84.

57. Noir, S.; Combes, M.-C.; Anthony, F.; Lashermes, P., Origin, diversity and evolution of NBS-type disease-resistance gene homologues in coffee trees (*Coffea* L.). *Molecular Genetics and Genomics* **2001,** *265* (4), 654-662.

58. Moffett, P., Chapter 1 - Mechanisms of recognition in dominant *R* gene mediated resistance. In *Advances in Virus Research*, Gad, L.; John, P. C., Eds. Academic Press: 2009; Vol. Volume 75, pp 1-229.

59. Zhou, T.; Wang, Y.; Chen, J. Q.; Araki, H.; Jing, Z.; Jiang, K.; Shen, J.; Tian, D., Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Molecular Genetics and Genomics* **2004,** *271*.

60. Alvarenga, S. M.; Caixeta, E. T.; Hufnagel, B.; Thiebaut, F.; Maciel-Zambolim, E.; Zambolimand, L.; Sakiyama, N. S., In silico identification of coffee genome expressed sequences potentially associated with resistance to diseases. *Genetics and molecular biology* **2010,** *33* (4), 795-806.

61. Friedman, A. R.; Baker, B. J., The evolution of resistance genes in multi-protein plant resistance systems. *Current Opinion in Genetics & Development* **2007,** *17* (6), 493-499.

62. Ribas, A. F.; Cenci, A.; Combes, M.-C.; Etienne, H.; Lashermes, P., Organization and molecular evolution of a disease-resistance gene cluster in coffee trees. *BMC Genomics* **2011,** *12* (1), 240.

63. Eulgem, T.; Somssich, I. E., Networks of WRKY transcription factors in defense signaling. *Current Opinion in Plant Biology* **2007,** *10* (4), 366-371.

64. Babu, M. M.; Luscombe, N. M.; Aravind, L.; Gerstein, M.; Teichmann, S. A., Structure and evolution of transcriptional regulatory networks. *Current Opinion in Structural Biology* **2004,** *14* (3), 283-291.

65. Cacas, J.-L.; Petitot, A.-S.; Bernier, L.; Estevan, J.; Conejero, G.; Mongrand, S.; Fernandez, D., Identification and characterization of the Non-race specific Disease Resistance 1 (NDR1) orthologous protein in coffee. *BMC Plant Biology* **2011,** *11* (1), 144.

66. Schluttenhofer, C.; Yuan, L., Regulation of specialized metabolism by WRKY transcription factors. *Plant Physiol.* **2015,** *167* (2), 295-306.

67. Gershenzon, J.; Dudareva, N., The function of terpene natural products in the natural world. *Nat Chem Biol* **2007,** *3* (7), 408-414.

68. Caarls, L.; Pieterse, C. M. J.; Van Wees, S. C. M., How salicylic acid takes transcriptional control over jasmonic acid signaling. *Frontiers in Plant Science* **2015,** *6* (170).

69. Navarro, L.; Bari, R.; Achard, P.; Lisón, P.; Nemri, A.; Harberd, N. P.; Jones, J. D. G., DELLAs control plant immune responses by modulating the balance of jasmonic acid and salicylic acid signaling. *Current Biology* **2008,** *18* (9), 650-655.

70. Windram, O.; Penfold, C. A.; Denby, K. J., Network modeling to understand plant immunity. *Annu. Rev. Phytopathol.* **2014,** *52,* 93-111.

71. Ramiro, D.; Jalloul, A.; Petitot, A. S.; Grossi-de-Sa, M. F.; Maluf, M.; Fernandez, D., Identification of coffee WRKY transcription factor genes and expression profiling in resistance responses to pathogens. *Tree Genet Genomes* **2010,** *6*.

72. Gil Azinheira, H.; do Céu Silva, M.; Talhinhas, P.; Medeira, C.; Maia, I.; Petitot, A.-S.; Fernandez, D., Non-host resistance responses of *Arabidopsis thaliana* to the coffee leaf rust fungus (*Hemileia vastatrix*). *Botany* **2010,** *88* (7), 621-629.

73. Lee, H.-A.; Lee, H.-Y.; Seo, E.; Lee, J.; Kim, S.-B.; Oh, S.; Choi, E.; Choi, E.; Lee, S. E.; Choi, D., Current understandings of plant nonhost resistance. *Molecular Plant-Microbe Interactions* **2016,** *30* (1), 5-15.

74. Ellis, J. G.; Lagudah, E. S.; Spielmeyer, W.; Dodds, P. N., The past, present and future of breeding rust resistant wheat. *Frontiers in Plant Science* **2014,** *5*, 641.

75. Vaast, P.; Bertrand, B.; Perriot, J.-J.; Guyot, B.; Génard, M., Fruit thinning and shade improve bean characteristics and beverage quality of coffee (*Coffea arabica* L.) under optimal conditions. *J. Sci. Food Agric.* **2006,** *86* (2), 197-204.

76. Kramer, D.; Breitenstein, B.; Kleinwächter, M.; Selmar, D., Stress metabolism in green coffee beans (*Coffea arabica* L.): expression of dehydrins and accumulation of GABA during drying. *Plant and Cell Physiology* **2010,** *51* (4), 546-553.

77. Joët, T.; Laffargue, A.; Salmona, J.; Doulbeau, S.; Descroix, F.; Bertrand, B.; de Kochko, A.; Dussert, S., Metabolic pathways in tropical dicotyledonous albuminous seeds: *Coffea arabica* as a case study. *New Phytologist* **2009,** *182* (1), 146-162.

78. De Castro, R. D.; Marraccini, P., Cytology, biochemistry and molecular changes during coffee fruit development. *Braz. J. Plant Physiol.* **2006,** *18*, 175-199.

79. Ivamoto, S. T.; Júnior, O. R.; Domingues, D. S.; dos Santos, T. B.; de Oliveira, F. F.; Pot, D.; Leroy, T.; Vieira, L. G. E.; Carazzolle, M. F.; Pereira, G. A. G., Transcriptome analysis of leaves, flowers and fruits perisperm of *Coffea arabica* L. reveals the differential expression of genes involved in raffinose biosynthesis. *PloS one* **2017,** *12* (1), e0169595.

80. Ságio, S. A.; Barreto, H. G.; Lima, A. A.; Moreira, R. O.; Rezende, P. M.; Paiva, L. V.; Chalfun-Junior, A., Identification and expression analysis of ethylene biosynthesis and signaling genes provides insights into the early and late coffee cultivars ripening pathway. *Planta* **2014,** *239* (5), 951-963.

81. Van de Poel, B.; Van Der Straeten, D., 1-aminocyclopropane-1-carboxylic acid (ACC) in plants: more than just the precursor of ethylene! *Frontiers in Plant Science* **2014,** *5*, 640.

82. Gaspari-Pezzopane, C. d.; Bonturi, N.; Guerreiro Filho, O.; Favarin, J. L.; Maluf, M. P., Gene expression profile during coffee fruit development and identification of candidate markers for phenological stages. *Pesquisa Agropecuária Brasileira* **2012,** *47*, 972-982.

83. Cação, S.; Leite, T.; Budzinski, I.; dos Santos, T.; Scholz, M.; Carpentieri-Pipolo, V.; Domingues, D.; Vieira, L.; Pereira, L., Gene expression and enzymatic activity of pectin methylesterase during fruit development and ripening in *Coffea arabica* L. *Genetics and Molecular Research* **2012,** *11* (3), 3186-3197.

84. Goulet, C.; Kamiyoshihara, Y.; Lam, Nghi B.; Richard, T.; Taylor, Mark G.; Tieman, Denise M.; Klee, Harry J., Divergence in the enzymatic activities of a tomato and *Solanum pennellii* alcohol acyltransferase impacts fruit volatile ester composition. *Molecular Plant* **2015,** *8* (1), 153-162.

85. Toledo, P. R.; Pezza, L.; Pezza, H. R.; Toci, A. T., Relationship between the different aspects related to coffee quality and their volatile compounds. *Compr. Rev. Food Sci. Food Saf.* **2016,** *15* (4), 705-719.

86. De Bruyn, F.; Zhang, S. J.; Pothakos, V.; Torres, J.; Lambot, C.; Moroni, A. V.; Callanan, M.; Sybesma, W.; Weckx, S.; De Vuyst, L., Exploring the impacts of postharvest processing on the microbiota and metabolite profiles during green coffee bean production. *Appl. Environ. Microbiol.* **2017,** *83* (1), e02398-16.

87. Leroy, T.; Ribeyre, F.; Bertrand, B.; Charmetant, P.; Dufour, M.; Montagnon, C.; Marraccini, P.; Pot, D., Genetics of coffee quality. *Braz. J. Plant Physiol.* **2006,** *18* (1), 229-242.

88. Folmer, B., *The craft and science of coffee*. Academic Press: 2017; Vol. 1, p 556.

89. Kwon, D.-J.; Jeong, H.-J.; Moon, H.; Kim, H.-N.; Cho, J.-H.; Lee, J.-E.; Hong, K. S.; Hong, Y.-S., Assessment of green coffee bean metabolites dependent on coffee quality using a 1 H NMR-based metabolomics approach. *Food Res. Int.* **2015,** *67*, 175-182.

90. Petisca, C.; Pérez-Palacios, T.; Farah, A.; Pinho, O.; Ferreira, I. M., Furans and other volatile compounds in ground roasted and espresso coffee using headspace solid-phase microextraction: Effect of roasting speed. *Food and Bioproducts Processing* **2013,** *91* (3), 233-241.

91. Joët, T.; Pot, D.; Ferreira, L. P.; Dussert, S.; Marraccini, P., Identification des déterminants moléculaires de la qualité du café par des approches de génomique fonctionnelle. Une revue. *Cahiers Agricultures* **2012,** *21* (2), 125-133.

92. Salmona, J.; Dussert, S.; Descroix, F.; De Kochko, A.; Bertrand, B.; Joët, T., Deciphering transcriptional networks that govern *Coffea arabica* seed development

using combined cDNA array and real-time RT-PCR approaches. *Plant molecular biology* **2008,** *66* (1-2), 105-124.

93. Privat, I.; Foucrier, S.; Prins, A.; Epalle, T.; Eychenne, M.; Kandalaft, L.; Caillet, V.; Lin, C.; Tanksley, S.; Foyer, C., Differential regulation of grain sucrose accumulation and metabolism in *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta) revealed through gene expression and enzyme activity analysis. *New Phytologist* **2008,** *178* (4), 781-797.

94. Ashihara, H.; Sano, H.; Crozier, A., Caffeine and related purine alkaloids: biosynthesis, catabolism, function and genetic engineering. *Phytochemistry* **2008,** *69* (4), 841-856.

95. Taguchi, H.; Sakaguchi, M.; Shimabayashi, Y., Trigonelline content in coffee beans and the thermal conversion of trigonelline into nicotinic acid during the roasting of coffee beans. *Agric. Biol. Chem.* **1985,** *49* (12), 3467-3471.

96. Moon, J.-K.; Yoo, H. S.; Shibamoto, T., Role of roasting conditions in the level of chlorogenic acid content in coffee beans: correlation with coffee acidity. *Journal of Agricultural and Food Chemistry* **2009,** *57* (12), 5365-5369.

97. Schenker, S.; Heinemann, C.; Huber, M.; Pompizzi, R.; Perren, R.; Escher, R., Impact of roasting conditions on the formation of aroma compounds in coffee beans. *Journal of food science* **2002,** *67* (1), 60-66.

98. Redgwell, R.; Fischer, M., Coffee carbohydrates. *Braz. J. Plant Physiol.* **2006,** *18* (1), 165-174.

99. Tran, H.; Lee, L. S.; Furtado, A.; Smyth, H.; Henry, R. J., Advances in genomics for the improvement of quality in coffee. *J. Sci. Food Agric.* **2016**.

100. Arya, M.; Rao, L. J. M., An impression of coffee carbohydrates. *Critical reviews in food science and nutrition* **2007,** *47* (1), 51-67.

101. Santos, J. R.; Sarraguça, M. C.; Rangel, A. O.; Lopes, J. A., Evaluation of green coffee beans quality using near infrared spectroscopy: A quantitative approach. *Food Chem.* **2012,** *135* (3), 1828-1835.

102. Cecilia, K.; Glaston, K.; Simon, M.; Renaud, B.; Fredrick, N., Volatile organic compounds in brewed Kenyan Arabica coffee genotypes by solid phase extraction gas chromatography mass spectrometry. *Food Science and Quality Management* **2012,** *8*, 18-26.

103. Ribeiro, J.; Augusto, F.; Salva, T.; Thomaziello, R.; Ferreira, M., Prediction of sensory properties of Brazilian Arabica roasted coffees by headspace solid phase microextraction-gas chromatography and partial least squares. *Anal. Chim. Acta* **2009,** *634* (2), 172-179.

104. Bertrand, B.; Etienne, H.; Lashermes, P.; Guyot, B.; Davrieux, F., Can near-infrared reflectance of green coffee be used to detect introgression in *Coffea arabica* cultivars? *J. Sci. Food Agric.* **2005,** *85* (6), 955-962.

105. Esquivel, P.; Jiménez, V. M., Functional properties of coffee and coffee by-products. *Food Res. Int.* **2012,** *46* (2), 488-495.

106. Eira, M. T.; Silva, E.; De Castro, R. D.; Dussert, S.; Walters, C.; Bewley, J. D.; Hilhorst, H. W., Coffee seed physiology. *Braz. J. Plant Physiol.* **2006,** *18* (1), 149-163.

107. Farah, A.; Donangelo, C. M., Phenolic compounds in coffee. *Braz. J. Plant Physiol.* **2006,** *18* (1), 23-36.

108. Chambers, E. M., R. World Coffee Research Sensory Lexicon. https://worldcoffeeresearch.org/media/documents/WCR_Sensory_Lexicon_Edition_1.1_2016.pdf (accessed 03/15/2016).

109. Consonni, R.; Cagliani, L. R.; Cogliati, C., NMR based geographical characterization of roasted coffee. *Talanta* **2012,** *88*, 420-426.

110. Flament, I.; Bessière-Thomas, Y., *Coffee flavor chemistry*. John Wiley & Sons: Chichester, England, 2002; p 424.

111. Bertrand, B.; Boulanger, R.; Dussert, S.; Ribeyre, F.; Berthiot, L.; Descroix, F.; Joët, T., Climatic factors directly impact the volatile organic compound fingerprint in green Arabica coffee bean as well as coffee beverage quality. *Food Chem.* **2012,** *135* (4), 2575-2583.

112. Etienne, H.; Anthony, F.; Dussert, S.; Fernandez, D.; Lashermes, P.; Bertrand, B., Biotechnological applications for the improvement of coffee (*Coffea arabica* L.). *In Vitro Cell Dev-Plant* **2002,** *38*.

113. Berthou, F.; Mathieu, C.; Vedel, F., Chloroplast and mitochondrial DNA variation as indicator of phylogenetic relationships in the genus *Coffea* L. *Theoretical and applied genetics* **1983,** *65* (1), 77-84.

114. Lashermes, P.; Cros, J.; Marmey, P.; Charrier, A., Use of random amplified DNA markers to analyse genetic variability and relationships of *Coffea* species. *Genetic resources and crop evolution* **1993,** *40* (2), 91-99.

115. Lashermes, P.; Combes, M.-C.; Robert, J.; Trouslot, P.; D'hont, A.; Anthony, F.; Charrier, A., Molecular characterisation and origin of the Coffea arabica L. genome. *Molecular and General Genetics MGG* **1999,** *261* (2), 259-266.

116. Lashermes, P.; Cros, J.; Combes, M. C.; Trouslot, P.; Anthony, F.; Hamon, S.; Charrier, A., Inheritance and restriction fragment length polymorphism of chloroplast DNA in the genus *Coffea* L. *Theoretical and Applied Genetics* **1996,** *93* (4), 626-632.

117. Moncada, P.; McCouch, S., Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. *Genome* **2004,** *47* (3), 501-509.

118. Fernandez, D.; Santos, P.; Agostini, C.; Bon, M. C.; Petitot, A. S.; C Silva, M.; Guerra-Guimarães, L.; Ribeiro, A.; Argout, X.; Nicole, M., Coffee (*Coffea arabica* L.) genes early expressed during infection by the rust fungus (*Hemileia vastatrix*). *Mol Plant Pathol* **2004,** *5*.

119. Lashermes, P.; Combes, M.-C.; Prakash, N.; Trouslot, P.; Lorieux, M.; Charrier, A., Genetic linkage map of *Coffea canephora*: effect of segregation distortion and analysis of recombination rate in male and female meioses. *Genome* **2001,** *44* (4), 589-595.

120. Aga, E.; Bekele, E.; Bryngelsson, T., Inter-simple sequence repeat (ISSR) variation in forest coffee trees (*Coffea arabica* L.) populations from Ethiopia. *Genetica* **2005,** *124* (2), 213-221.

121. Poncet, V.; Rondeau, M.; Tranchant, C.; Cayrel, A.; Hamon, S.; de Kochko, A.; Hamon, P., SSR mining in coffee tree EST databases: potential use of EST–SSRs as

markers for the Coffea genus. *Molecular Genetics and Genomics* **2006,** *276* (5), 436-449.

122. Noir, S.; Patheyron, S.; Combes, M.-C.; Lashermes, P.; Chalhoub, B., Construction and characterisation of a BAC library for genome analysis of the allotetraploid coffee species (Coffea arabica L.). *Theoretical and Applied Genetics* **2004,** *109* (1), 225-230.

123. Dereeper, A.; Guyot, R.; Tranchant-Dubreuil, C.; Anthony, F.; Argout, X.; de Bellis, F.; Combes, M.-C.; Gavory, F.; de Kochko, A.; Kudrna, D.; Leroy, T.; Poulain, J.; Rondeau, M.; Song, X.; Wing, R.; Lashermes, P., BAC-end sequences analysis provides first insights into coffee (*Coffea canephora* P.) genome composition and evolution. *Plant Molecular Biology* **2013,** *83* (3), 177-189.

124. Samson, N.; Bausher, M. G.; Lee, S.-B.; Jansen, R. K.; Daniell, H., The complete nucleotide sequence of the coffee (Coffea arabica L.) chloroplast genome: organization and implications for biotechnology and phylogenetic relationships amongst angiosperms. *Plant Biotechnology Journal* **2007,** *5* (2), 339-353.

125. Lopes, F. R.; Jjingo, D.; da Silva, C. R. M.; Andrade, A. C.; Marraccini, P.; Teixeira, J. B.; Carazzolle, M. F.; Pereira, G. A. G.; Pereira, L. F. P.; Vanzela, A. L. L.; Wang, L.; Jordan, I. K.; Carareto, C. M. A., Transcriptional Activity, Chromosomal Distribution and Expression Effects of Transposable Elements in Coffea Genomes. *PLOS ONE* **2013,** *8* (11), e78931.

126. Combes, M.-C.; Dereeper, A.; Severac, D.; Bertrand, B.; Lashermes, P., Contribution of subgenomes to the transcriptome and their intertwined regulation in the

allopolyploid Coffea arabica grown at contrasted temperatures. *New Phytologist* **2013,** *200* (1), 251-260.

127. Paschoal, A.; Fernandes, E.; Silva, J.; Lopes, F.; Pereira, L.; Domingues, D., CoffeebEST: an integrated resource for *Coffea* spp expressed sequence tags. *Genetics and Molecular Research* **2014,** *13* (4), 10913-10920.

128. Dereeper, A.; Bocs, S.; Rouard, M.; Guignon, V.; Ravel, S.; Tranchant-Dubreuil, C.; Poncet, V.; Garsmeur, O.; Lashermes, P.; Droc, G., The coffee genome hub: a resource for coffee genomes. *Nucleic acids research* **2015,** *43* (D1), D1028-D1035.

129. Andrade, A. C.; Carneiro, D. A.; Da Silva Jr, O. B.; Marraccini, P.; Grattapaglia, D., Towards GWAS and genomic prediction in coffee: development and validation of a 26K SNP chip for *Coffea canephora*.[W173]. **2017**.

130. Zeng, D.; Tian, Z.; Rao, Y.; Dong, G.; Yang, Y.; Huang, L.; Leng, Y.; Xu, J.; Sun, C.; Zhang, G., Rational design of high-yield and superior-quality rice. *Nature Plants* **2017,** *3*, 17031.

131. Yang, Q.; Balint-Kurti, P.; Xu, M., Quantitative disease resistance: dissection and adoption in maize. *Molecular Plant* **2017,** *10* (3), 402-413.

132. Mahan, A. L.; Murray, S. C.; Crosby, K.; Scott, M. P., Quality protein maize germplasm characterized for amino acid profiles and endosperm opacity. *Crop Sci.* **2014,** *54* (3), 863-872.

133. de Carvalho, K.; Petkowicz, C. L.; Nagashima, G. T.; Bespalhok Filho, J. C.; Vieira, L. G.; Pereira, L. F.; Domingues, D. S., Homeologous genes involved in

mannitol synthesis reveal unequal contributions in response to abiotic stress in Coffea arabica. *Molecular Genetics and Genomics* **2014,** *289* (5), 951-963.

134. Denoeud, F.; Carretero-Paulet, L.; Dereeper, A.; Droc, G.; Guyot, R.; Pietrella, M.; Zheng, C.; Alberti, A.; Anthony, F.; Aprea, G.; Aury, J.-M.; Bento, P.; Bernard, M.; Bocs, S.; Campa, C.; Cenci, A.; Combes, M.-C.; Crouzillat, D.; Da Silva, C.; Daddiego, L.; De Bellis, F.; Dussert, S.; Garsmeur, O.; Gayraud, T.; Guignon, V.; Jahn, K.; Jamilloux, V.; Joët, T.; Labadie, K.; Lan, T.; Leclercq, J.; Lepelley, M.; Leroy, T.; Li, L.-T.; Librado, P.; Lopez, L.; Muñoz, A.; Noel, B.; Pallavicini, A.; Perrotta, G.; Poncet, V.; Pot, D.; Priyono; Rigoreau, M.; Rouard, M.; Rozas, J.; Tranchant-Dubreuil, C.; VanBuren, R.; Zhang, Q.; Andrade, A. C.; Argout, X.; Bertrand, B.; de Kochko, A.; Graziosi, G.; Henry, R. J.; Jayarama; Ming, R.; Nagai, C.; Rounsley, S.; Sankoff, D.; Giuliano, G.; Albert, V. A.; Wincker, P.; Lashermes, P., The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* **2014,** *345* (6201), 1181-1184.

135. Echeverría Beirute, F.; Barquero Miranda, M.; Gómez Alpízar, L., Evaluación de la susceptibilidad de genotipos de café (*Coffea* spp) al hongo *Mycena citricolor* y determinación de genes candidatos de defensa. *Boletín Promecafé* 2013, p 3.

136. Dugas, D. V.; Monaco, M. K.; Olson, A.; Klein, R. R.; Kumari, S.; Ware, D.; Klein, P. E., Functional annotation of the transcriptome of *Sorghum bicolor* in response to osmotic stress and abscisic acid. *BMC Genomics* **2011,** *12* (1), 1.

137. Bian, Y.; Holland, J. B., Enhancing genomic prediction with genome-wide association studies in multiparental maize populations. *Heredity* **2017**.

138. Mackay, T. F. C., Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat Rev Genet* **2014,** *15* (1), 22-33.

139. Crick, F., Central dogma of molecular biology. *Nature* **1970,** *227* (5258), 561-563.

140. Shapiro, J. A., Revisiting the central dogma in the 21st century. *Ann. N. Y. Acad. Sci.* **2009,** *1178* (1), 6-28.

141. Shapiro, J. A., Physiology of the read–write genome. *The Journal of Physiology* **2014,** *592* (11), 2319-2341.

142. Alvarez, M.; Schrey, A. W.; Richards, C. L., Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? *Molecular ecology* **2015,** *24* (4), 710-725.

143. Olson, A.; Klein, R. R.; Dugas, D. V.; Lu, Z.; Regulski, M.; Klein, P. E.; Ware, D., Expanding and vetting gene annotations through transcriptome and methylome sequencing. *The Plant Genome* **2014,** *7* (2).

144. Kranner, I.; Minibayeva, F. V.; Beckett, R. P.; Seal, C. E., What is stress? Concepts, definitions and applications in seed science. *New Phytologist* **2010,** *188* (3), 655-673.

145. Cristancho, M. A.; Botero-Rozo, D. O.; Giraldo, W.; Tabima, J.; Riaño-Pachón, D. M.; Escobar, C.; Rozo, Y.; Rivera, L. F.; Durán, A.; Restrepo, S., Annotation of a hybrid partial genome of the coffee rust (*Hemileia vastatrix*) contributes to the gene repertoire catalog of the Pucciniales. *Frontiers in plant science* **2014,** *5*, 594.

146. Majewski, J.; Pastinen, T., The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends in Genetics* **2011,** *27* (2), 72-79.

147. Gilad, Y.; Rifkin, S. A.; Pritchard, J. K., Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends in Genetics* **2008,** *24* (8), 408-415.

148. Medrano, J.; Deynze, A.; Cantu, D.; Hulse-Kemp, A., *Coffee arabica* UCDv0.5. **2017**.

149. Bertrand, B.; Bardil, A.; Baraille, H.; Dussert, S.; Doulbeau, S.; Dubois, E.; Severac, D.; Dereeper, A.; Etienne, H., The greater phenotypic homeostasis of the allopolyploid *Coffea arabica* improved the transcriptional homeostasis over that of both diploid parents. *Plant and Cell Physiology* **2015,** *56* (10), 2035-2051.

150. Hailemichael, M. Influence of genotype, location and processing methods on the quality of coffee (*Coffea arabica* L.). Thesis, Hawassa University, Ethiopia, 2009.

151. Läderach, P.; Oberthür, T.; Cook, S.; Estrada Iza, M.; Pohlan, J. A.; Fisher, M.; Rosales Lechuga, R., Systematic agronomic farm management for improved coffee quality. *Field Crops Research* **2011,** *120* (3), 321-329.

152. DaMatta, F. M.; Ronchi, C. P.; Maestri, M.; Barros, R. S., Ecophysiology of coffee growth and production. *Brazilian Journal of Plant Physiology* **2007,** *19*, 485-510.

153. Bote, A. D.; Jan, V., Branch growth dynamics, photosynthesis, yield and bean size distribution in response to fruit load manipulation in coffee trees. *Trees* **2016,** *30* (4), 1275-1285.

154. Waller, J. M.; Bigger, M.; Hillocks, R. J., *Coffee pests, diseases and their management*. CABI: Egham, Surrey, UK, 2007; p 400.

155. de Melo Virginio Filho, E.; Domian, C. A., *Prevención y control de la roya del café*. Centro Agronómico Tropical de Investigación y Enseñanza (CATIE): Turrialba, Costa Rica, 2015; p 98.

156. Buffo, R. A.; Cardelli-Freire, C., Coffee flavour: an overview. *Flavour Fragrance J.* **2004,** *19* (2), 99-104.

157. Fayle, S. E.; Gerrard, J. A., *The Maillard Reaction*. Royal Society of Chemistry: Norwich, UK 2002; Vol. 5, p 117.

158. SCAA Cupping Protocols. http://scaa.org/?page=resources&d=cupping-protocols (accessed 11/29/2016).

159. Chambers, E.; Sanchez, K.; Phan, U. X. T.; Miller, R.; Civille, G. V.; Di Donfrancesco, B., Development of a "living" lexicon for descriptive sensory analysis of brewed coffee. *J. Sens. Stud.* **2016,** *31* (6), 465-480.

160. Kerth, C., Determination of volatile aroma compounds in beef using differences in steak thickness and cook surface temperature. *Meat Sci.* **2016,** *117*, 27-35.

161. McCulloch, C. E., Generalized linear models. *Journal of the American Statistical Association* **2000,** *95* (452), 1320-1324.

162. Siegel, M., Sterol-inhibiting fungicides: effects on sterol biosynthesis and sites of action. *Plant Diseases* **1981,** *65* (12), 986-989.

163. ICAFE Costos de produccion agricola de cafe fruta. http://www.icafe.cr/wp-content/uploads/informacion_mercado/costos_actividad/produccion/CPACBaja1516.pdf (accessed 3/15/17).

164. Roy, B.; Kirchner, J., Evolutionary dynamics of pathogen resistance and tolerance. *Evolution* **2000,** *54* (1), 51-63.

165. de Oliveira Fassio, L.; Malta, M. R.; Carvalho, G. R.; Liska, G. R.; de Lima, P. M.; Pimenta, C. J., Sensory description of cultivars (*Coffea arabica* L.) resistant to rust and its correlation with caffeine, trigonelline, and chlorogenic acid compounds. *Beverages* **2016,** *2* (1), 1.

166. Brune, K.; Renner, B.; Tiegs, G., Acetaminophen/paracetamol: A history of errors, failures and false decisions. *Eur. J. Pain* **2015,** *19* (7), 953-965.

167. Mauch-Mani, B.; Slusarenko, A. J., Production of salicylic acid precursors is a major function of phenylalanine ammonia-lyase in the resistance of *Arabidopsis* to *Peronospora parasitica*. *Plant Cell* **1996,** *8* (2), 203-212.

168. Herrmann, K. M., The Shikimate Pathway: early steps in the biosynthesis of aromatic compounds. *Plant Cell* **1995,** *7* (7), 907-919.

169. Daykin, T. Dissecting the indigo pathway. RMIT University, 2011.

170. Browning, D.; Moayyad, S., Chapter 5 - Social Sustainability—Community, Livelihood, and Tradition A2 - Folmer, Britta. In *The Craft and Science of Coffee*, Academic Press: 2017; pp 109-131.

171. Brain, S. Coffee drinking statistics. http://www.statisticbrain.com/coffee-drinking-statistics/ (accessed 7/19/2017).

172. ICO, Historical Data on the Global Coffee Trade. **2016**.

173. McCook, S., Global rust belt: *Hemileia vastatrix* and the ecological integration of world coffee production since 1850. *Journal of Global History* **2006,** *1* (2), 177-195.

174. Joët, T.; Laffargue, A.; Salmona, J.; Doulbeau, S.; Descroix, F.; Bertrand, B.; de Kochko, A.; Dussert, S., Metabolic pathways in tropical dicotyledonous albuminous seeds: Coffea arabica as a case study. *The New Phytologist* **2009,** *182* (1), 146-162.

175. Speer, K.; Kölling-Speer, I., The lipid fraction of the coffee bean. *Braz. J. Plant Physiol.* **2006,** *18*, 201-216.

176. Cagliari, A.; Margis, R.; dos Santos Maraschin, F.; Turchetto-Zolet, A. C.; Loss, G.; Margis-Pinheiro, M., Biosynthesis of triacylglycerols (TAGs) in plants and algae. *International Journal of Plant Biology* **2011,** *2* (1), e10.

177. Pavesi Arisseto, A.; Vicente, E.; Soares Ueno, M.; Verdiani Tfouni, S. A. l.; De Figueiredo Toledo, M. C. l., Furan levels in coffee as influenced by species, roast degree, and brewing procedures. *Journal of agricultural and food chemistry* **2011,** *59* (7), 3118-3124.

178. Hurtado-Benavides, A.; Dorado A, D.; Sánchez-Camargo, A. d. P., Study of the fatty acid profile and the aroma composition of oil obtained from roasted Colombian coffee beans by supercritical fluid extraction. *The Journal of Supercritical Fluids* **2016,** *113* (Supplement C), 44-52.

179. Figueiredo, L. P.; Borem, F. M.; Ribeiro, F. C.; Giomo, G. S.; da Silva Taveira, J. H.; Malta, M. R., Fatty acid profiles and parameters of quality of specialty coffees produced in different Brazilian regions. *African Journal of Agricultural Research* **2015,** *10* (35), 3484-3493.

180. Echeverria-Beirute, F.; Murray, S. C.; Klein, P.; Kerth, C.; Miller, R.; Bertrand, B., Rust and thinning management effect on cup quality and plant performance for two cultivars of *Coffea arabica* L. *Journal of Agricultural and Food Chemistry* **2017**, 12.

181. Zhong, S.; Joung, J.-G.; Zheng, Y.; Chen, Y.; Liu, B.; Shao, Y.; Xiang, J. Z.; Fei, Z.; Giovannoni, J. J., High-throughput illumina strand-specific RNA sequencing library preparation. *Cold Spring Harb Protoc* **2011,** *2011* (8), 940-9.

182. Andrews, S. *FastQC A quality control tool for high throughput sequence data*, Babraham Bioinformatics: http://www.bioinformatics.babraham.ac.uk/projects/fastqc, 2010.

183. Bolger, A. M.; Lohse, M.; Usadel, B., Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, btu170.

184. Yoon, S.; Nam, D., Gene dispersion is the key determinant of the read count bias in differential expression analysis of RNA-seq data. *BMC Genomics* **2017,** *18* (1), 408.

185. Bullard, J. H.; Purdom, E.; Hansen, K. D.; Dudoit, S., Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinf.* **2010,** *11* (1), 94.

186. Westerhuis, J. A.; Kourti, T.; MacGregor, J. F., Analysis of multiblock and hierarchical PCA and PLS models. *Journal of chemometrics* **1998,** *12* (5), 301-321.

187. Lever, J.; Krzywinski, M.; Altman, N., Principal component analysis. *Nature Methods* **2017,** *14*, 641.

188. Tian, T.; Liu, Y.; Yan, H.; You, Q.; Yi, X.; Du, Z.; Xu, W.; Su, Z., agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Research* **2017,** *45* (W1), W122-W129.

189. Linstrom, P.; Mallard, W., NIST chemistry webbook, NIST standard reference database number 69, National Institute of Standards and Technology, Gaithersburg MD, 20899. 2010.

190. Erten, E. S.; Cadwallader, K. R., Identification of predominant aroma components of raw, dry roasted and oil roasted almonds. *Food Chem.* **2017,** *217* (Supplement C), 244-253.

191. Yashin, A.; Yashin, Y.; Xia, X.; Nemzer, B., Chromatographic methods for coffee analysis: a review. *2017* **2017,** *6* (4).

192. Sousa, C. A. F. d.; Sodek, L., The metabolic response of plants to oxygen deficiency. *Braz. J. Plant Physiol.* **2002,** *14*, 83-94.

193. Baggenstoss, J.; Poisson, L.; Kaegi, R.; Perren, R.; Escher, F., Coffee roasting and aroma formation: application of different time−temperature conditions. *Journal of Agricultural and Food Chemistry* **2008,** *56* (14), 5836-5846.

194. Ginz, M.; Balzer, H. H.; Bradbury, A. G.; Maier, H. G., Formation of aliphatic acids by carbohydrate degradation during roasting of coffee. *European Food Research and Technology* **2000,** *211* (6), 404-410.

195. Vickers, C. E.; Gershenzon, J.; Lerdau, M. T.; Loreto, F., A unified mechanism of action for volatile isoprenoids in plant abiotic stress. *Nature Chemical Biology* **2009,** *5*, 283.

196. Del Terra, L.; Lonzarich, V.; Asquini, E.; Navarini, L.; Graziosi, G.; Suggi Liverani, F.; Pallavicini, A., Functional characterization of three *Coffea arabica* L. monoterpene synthases: Insights into the enzymatic machinery of coffee aroma. *Phytochemistry* **2013,** *89* (Supplement C), 6-14.

197. Baxter, A.; Mittler, R.; Suzuki, N., ROS as key players in plant stress signalling. *Journal of Experimental Botany* **2014,** *65* (5), 1229-1240.

198. Pinot, F.; Beisson, F., Cytochrome P450 metabolizing fatty acids in plants: characterization and physiological roles. *FEBS Journal* **2011,** *278* (2), 195-205.

199. Chen, G.; Liu, C.; Gao, Z.; Zhang, Y.; Jiang, H.; Zhu, L.; Ren, D.; Yu, L.; Xu, G.; Qian, Q., OsHAK1, a high-affinity potassium transporter, positively regulates responses to drought stress in rice. *Frontiers in Plant Science* **2017,** *8* (1885).

200. Lee, S.-B.; Jung, S.-J.; Go, Y.-S.; Kim, H.-U.; Kim, J.-K.; Cho, H.-J.; Park, O. K.; Suh, M.-C., Two *Arabidopsis* 3-ketoacyl CoA synthase genes, *KCS20* and *KCS2/DAISY*, are functionally redundant in cuticular wax and root suberin biosynthesis, but differentially controlled by osmotic stress. *The Plant Journal* **2009,** *60* (3), 462-475.

201. Florez, J. C.; Mofatto, L. S.; do Livramento Freitas-Lopes, R.; Ferreira, S. S.; Zambolim, E. M.; Carazzolle, M. F.; Zambolim, L.; Caixeta, E. T., High throughput transcriptome analysis of coffee reveals prehaustorial resistance in response to *Hemileia vastatrix* infection. *Plant Molecular Biology* **2017,** *95* (6), 607-623.

202. Chepyshko, H.; Lai, C.-P.; Huang, L.-M.; Liu, J.-H.; Shaw, J.-F., Multifunctionality and diversity of GDSL esterase/lipase gene family in rice (*Oryza*

sativa L. japonica) genome: new insights from bioinformatics analysis. *BMC Genomics* **2012,** *13* (1), 309.

203. Dadakova, K.; Havelkova, M.; Kurkova, B.; Tlolkova, I.; Kasparovsky, T.; Zdrahal, Z.; Lochman, J., Proteome and transcript analysis of *Vitis vinifera* cell cultures subjected to *Botrytis cinerea* infection. *Journal of Proteomics* **2015,** *119*, 143-153.

204. Chebli, Y.; Geitmann, A., Cellular growth in plants requires regulation of cell wall biochemistry. *Current Opinion in Cell Biology* **2017,** *44*, 28-35.

205. Sasidharan, R.; Voesenek, L. A. C. J.; Pierik, R., Cell wall modifying proteins mediate plant acclimatization to biotic and abiotic stresses. *Critical Reviews in Plant Sciences* **2011,** *30* (6), 548-562.

206. Ziegler, J.; Facchini, P. J., Alkaloid biosynthesis: metabolism and trafficking. *Annu. Rev. Plant Biol.* **2008,** *59* (1), 735-769.

207. Bai, Y.; Dougherty, L.; Li, M.; Fazio, G.; Cheng, L.; Xu, K., A natural mutation-led truncation in one of the two aluminum-activated malate transporter-like genes at the *Ma* locus is associated with low fruit acidity in apple. *Molecular Genetics and Genomics* **2012,** *287* (8), 663-678.

208. Fernie, A. R.; Carrari, F.; Sweetlove, L. J., Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. *Current Opinion in Plant Biology* **2004,** *7* (3), 254-261.

209. Noble, W. S., How does multiple testing correction work? *Nature Biotechnology* **2009,** *27*, 1135.

210. Houston, K.; Tucker, M. R.; Chowdhury, J.; Shirley, N.; Little, A., The plant cell wall: a complex and dynamic structure as revealed by the responses of genes under stress conditions. *Frontiers in Plant Science* **2016,** *7* (984).

211. Le Gall, H.; Philippe, F.; Domon, J.-M.; Gillet, F.; Pelloux, J.; Rayon, C., Cell wall metabolism in response to abiotic stress. *Plants* **2015,** *4* (1), 112.

212. Niinemets, Ü.; Kännaste, A.; Copolovici, L., Quantitative patterns between plant volatile emissions induced by biotic stresses and the degree of damage. *Frontiers in Plant Science* **2013,** *4*, 262.

213. Yasuda, E.; Ebinuma, H.; Wabiko, H., A novel glycine-rich/hydrophobic 16 kDa polypeptide gene from tobacco: similarity to proline-rich protein genes and its wound-inducible and developmentally regulated expression. *Plant Molecular Biology* **1997,** *33* (4), 667-678.

214. Qin, L.-X.; Zhang, D.-J.; Huang, G.-Q.; Li, L.; Li, J.; Gong, S.-Y.; Li, X.-B.; Xu, W.-L., Cotton GhHyPRP3 encoding a hybrid proline-rich protein is stress inducible and its overexpression in *Arabidopsis* enhances germination under cold temperature and high salinity stress conditions. *Acta Physiologiae Plantarum* **2013,** *35* (5), 1531-1542.

215. Farn, J. L.; Strugnell, R. A.; Hoyne, P. A.; Michalski, W. P.; Tennent, J. M., Molecular characterization of a secreted enzyme with phospholipase B activity from *Moraxella bovis*. *Journal of bacteriology* **2001,** *183* (22), 6717-6720.

216. Zhang, J.-h.; Sun, H.-l.; Chen, S.-y.; Zeng, L.; Wang, T.-t., Anti-fungal activity, mechanism studies on α-phellandrene and nonanal against *Penicillium cyclopium*. *Botanical Studies* **2017,** *58* (1), 13.

217. Tsitsigiannis, D. I.; Keller, N. P., Oxylipins as developmental and host–fungal communication signals. *Trends in Microbiology* **2007,** *15* (3), 109-118.

218. Villarreal, D.; Laffargue, A.; Posada, H.; Bertrand, B.; Lashermes, P.; Dussert, S., Genotypic and environmental effects on coffee (*Coffea arabica* L.) bean fatty acid profile: impact on variety and origin chemometric determination. *Journal of Agricultural and Food Chemistry* **2009,** *57* (23), 11321-11327.

219. Lim, G.-H.; Singhal, R.; Kachroo, A.; Kachroo, P., Fatty acid– and lipid-mediated signaling in plant defense. *Annu. Rev. Phytopathol.* **2017,** *55* (1), 505-536.

220. Noctor, G., Metabolic signalling in defence and stress: the central roles of soluble redox couples. *Plant, Cell & Environment* **2006,** *29* (3), 409-425.

221. Diniz, I.; Talhinhas, P.; Azinheira, H.; Varzea, V.; Medeira, C.; Maia, I.; Petitot, A.; Nicole, M.; Fernandez, D.; Silva, M., Cellular and molecular analyses of coffee resistance to Hemileia vastatrix and nonhost resistance to Uromyces vignae in the resistance-donor genotype HDT832/2. *Eur J Plant Pathol* **2012,** *133*.

222. DaMatta, F. M.; Ronchi, C. P.; Maestri, M.; Barros, R. S., Ecophysiology of coffee growth and production. *Braz. J. Plant Physiol.* **2007,** *19* (4), 485-510.

223. Silva, M. C.; Nicole, M.; Guerra-Guimarães, L.; Rodrigues, C. J., Hypersensitive cell death and post-haustorial defense responses arrest the orange rust (*Hemileia vastatrix*) growth in resistant coffee leaves. *Physiol Mol Plant Pathol* **2002,** *60*.

224. Serrano, I.; Romero-Puertas, M. C.; Sandalio, L. M.; Olmedilla, A., The role of reactive oxygen species and nitric oxide in programmed cell death associated with self-incompatibility. *Journal of experimental botany* **2015,** *66* (10), 2869-2876.

225. Xiong, L.; Yang, Y., Disease resistance and abiotic stress tolerance in rice are inversely modulated by an Abscisic Acid–Inducible Mitogen-Activated Protein Kinase. *Plant Cell* **2003,** *15* (3), 745-759.

226. Oh, E.-S.; Lee, Y.; Chae, W.; Rameneni, J.; Park, Y.-S.; Lim, Y.; Oh, M.-H., Biochemical analysis of the role of Leucine-Rich Repeat Receptor-like Kinases and the Carboxy-Terminus of Receptor Kinases in regulating kinase activity in *Arabidopsis thaliana* and *Brassica oleracea*. *Molecules* **2018,** *23* (1), 236.

227. Monteiro, A. C. A.; Resende, M. L. V.; Valente, T. C. T.; Ribeiro Junior, P. M.; Pereira, V. F.; Costa, J. R.; Silva, J. A. G., Manganese phosphite in coffee defence against *Hemileia vastatrix*, the coffee rust fungus: Biochemical and molecular analyses. *Journal of Phytopathology* **2016,** *164* (11-12), 1043-1053.

228. Menzel, M.; Sletvold, N.; Ågren, J.; Hansson, B., Inbreeding affects gene expression differently in two self-incompatible *Arabidopsis lyrata* populations with similar levels of inbreeding depression. *Molecular Biology and Evolution* **2015,** *32* (8), 2036-2047.

229. Birch, P. R. J.; Avrova, A. O.; Duncan, J. M.; Lyon, G. D.; Toth, R. L., Isolation of potato genes that are induced during an early stage of the hypersensitive response to *Phytophthora infestans*. *Molecular Plant-Microbe Interactions* **1999,** *12* (4), 356-361.

230. Nersissian, A. M.; Valentine, J. S.; Immoos, C.; Hill, M. G.; Hart, P. J.; Williams, G.; Herrmann, R. G., Uclacyanins, stellacyanins, and plantacyanins are distinct subfamilies of phytocyanins: Plant-specific mononuclear blue copper proteins. *Protein Science* **1998,** *7* (9), 1915-1929.

231. Munné-Bosch, S., The role of α-tocopherol in plant stress tolerance. *Journal of Plant Physiology* **2005,** *162* (7), 743-748.

232. Lee, E.-J.; Facchini, P. J., Tyrosine aminotransferase contributes to benzylisoquinoline alkaloid biosynthesis in Opium Poppy. *Plant Physiol.* **2011,** *157* (3), 1067-1078.

233. Guidetti-Gonzalez, S.; Freitas-Astúa, J.; Amaral, A. M. d.; Martins, N. F.; Mehta, A.; Silva, M. S.; Carrer, H., Genes associated with hypersensitive response (HR) in the citrus EST database (CitEST). *Genetics and Molecular Biology* **2007,** *30*, 943-956.

234. Balmant, K. M.; Parker, J.; Yoo, M.-J.; Zhu, N.; Dufresne, C.; Chen, S., Redox proteomics of tomato in response to *Pseudomonas syringae* infection. *Horticulture Research* **2015,** *2*, 15043.

235. Dong, Y.; Deng, M.; Zhao, Z.; Fan, G., Quantitative proteomic and transcriptomic study on autotetraploid *Paulownia* and its diploid parent reveal key metabolic processes associated with *Paulownia* autotetraploidization. *Frontiers in Plant Science* **2016,** *7*, 892.

236. Moeder, W.; Barry, C. S.; Tauriainen, A. A.; Betz, C.; Tuomainen, J.; Utriainen, M.; Grierson, D.; Sandermann, H.; Langebartels, C.; Kangasjärvi, J., Ethylene synthesis regulated by biphasic induction of 1-aminocyclopropane-1-carboxylic acid synthase and 1-aminocyclopropane-1-carboxylic acid oxidase genes is required for hydrogen peroxide accumulation and cell death in ozone-exposed tomato. *Plant Physiol.* **2002,** *130* (4), 1918-1926.

237. Guerra-Guimarães, L.; Silva, M. C.; Struck, C.; Loureiro, A.; Nicole, M.; Rodrigues, C. J.; Ricardo, C. P. P., Chitinases of *Coffea arabica* genotypes resistant to orange rust *Hemileia vastatrix*. *Biol. Plant.* **2009,** *53* (4), 702.

238. Guzzo, S. D.; Harakava, R.; Tsai, S. M., Identification of coffee genes expressed during systemic acquired resistance and incompatible interaction with *Hemileia vastatrix*. *J Phytopathol* **2009,** *157*.

239. Balestrini, R.; Salvioli, A.; Dal Molin, A.; Novero, M.; Gabelli, G.; Paparelli, E.; Marroni, F.; Bonfante, P., Impact of an arbuscular mycorrhizal fungus versus a mixed microbial inoculum on the transcriptome reprogramming of grapevine roots. *Mycorrhiza* **2017,** *27* (5), 417-430.

240. Kottapalli, K. R.; Sarla, N.; Kikuchi, S., *In silico* insight into two rice chromosomal regions associated with submergence tolerance and resistance to bacterial leaf blight and gall midge. *Biotechnology Advances* **2006,** *24* (6), 561-589.

241. Li, J.; Su, X.; Wang, Y.; Yang, W.; Pan, Y.; Su, C.; Zhang, X., Genome-wide identification and expression analysis of the BTB domain-containing protein gene family in tomato. *Genes & Genomics* **2018,** *40* (1), 1-15.

242. Tognetti, V. B.; Van Aken, O.; Morreel, K.; Vandenbroucke, K.; van de Cotte, B.; De Clercq, I.; Chiwocha, S.; Fenske, R.; Prinsen, E.; Boerjan, W.; Genty, B.; Stubbs, K. A.; Inzé, D.; Van Breusegem, F., Perturbation of Indole-3-Butyric Acid homeostasis by the UDP-Glucosyltransferase *UGT74E2* modulates *Arabidopsis* architecture and water stress tolerance. *Plant Cell* **2010,** *22* (8), 2660-2679.

243. Belmonte, M.; Elhiti, M.; Ashihara, H.; Stasolla, C., Brassinolide-improved development of *Brassica napus* microspore-derived embryos is associated with increased activities of purine and pyrimidine salvage pathways. *Planta* **2011,** *233* (1), 95-107.

244. Yang, L.; Qin, L.; Liu, G.; Peremyslov, V. V.; Dolja, V. V.; Wei, Y., Myosins XI modulate host cellular responses and penetration resistance to fungal pathogens. *Proceedings of the National Academy of Sciences* **2014,** *111* (38), 13996-14001.

245. Phipps, K. R.; Charette, J. M.; Baserga, S. J., The small subunit processome in ribosome biogenesis—progress and prospects. *Wiley Interdisciplinary Reviews: RNA* **2011,** *2* (1), 1-21.

246. Dolan, W. L.; Chapple, C., Conservation and divergence of mediator structure and function: insights from plants. *Plant and Cell Physiology* **2017,** *58* (1), 04-21.

247. Baena-González, E., Energy signaling in the regulation of gene expression during stress. *Molecular Plant* **2010,** *3* (2), 300-313.

248. Margaria, P.; Abbà, S.; Palmano, S., Novel aspects of grapevine response to phytoplasma infection investigated by a proteomic and phospho-proteomic approach with data integration into functional networks. *BMC Genomics* **2013,** *14* (1), 38.

249. Zhao, Y.; Wang, Y.; Yang, H.; Wang, W.; Wu, J.; Hu, X., Quantitative proteomic analyses identify ABA-related proteins and signal pathways in maize leaves under drought conditions. *Frontiers in Plant Science* **2016,** *7*, 1827.

APPENDIX 1

Summary statistics of the SCAA attributes found to vary with the treatments.

SCAA attribute	Cultivar	Inbred				Hybrid			
	Rust control	Yes	No	Yes	No	Yes	No	Yes	No
	Fruit thinning	0%	0%	50%	50%	0%	0%	50%	50%
Clean cup	Mean	6.08^{ab}	6.21^{a}	6.00^{ab}	6.04^{ab}	5.75^{b}	6.17^{a}	6.13^{a}	6.29^{a}
	Std Dev	0.10	0.32	0.27	0.08	0.50	0.19	0.16	0.16
	Min	6.00	6.00	5.67	6.00	5.00	6.00	6.00	6.17
	Max	6.17	6.67	6.33	6.17	6.00	6.33	6.33	6.50
Acidity	Mean	6.46^{ab}	6.50^{ab}	6.83^{a}	6.58^{ab}	6.13^{ab}	6.54^{ab}	6.50^{ab}	6.79^{a}
	Std Dev	0.32	0.49	0.24	0.40	0.60	0.37	0.59	0.25
	Min	6.00	6.00	6.50	6.17	5.67	6.17	5.67	6.50
	Max	6.67	7.00	7.00	7.00	7.00	7.00	7.00	7.00
Body	Mean	6.17^{a}	5.96^{ab}	6.17^{a}	6.13^{a}	5.83^{b}	6.08^{a}	5.96^{ab}	6.04^{ab}
	Std Dev	0.24	0.08	0.14	0.16	0.19	0.17	0.08	0.08
	Min	6.00	5.83	6.00	6.00	5.67	6.00	5.83	6.00
	Max	6.50	6.00	6.33	6.33	6.00	6.33	6.00	6.17
Overall	Mean	6.13^{ab}	6.29^{a}	6.00^{ab}	6.21^{ab}	5.71^{b}	6.25^{ab}	6.17^{ab}	6.42^{a}
	Std Dev	0.16	0.21	0.54	0.34	0.67	0.29	0.14	0.42
	Min	6.00	6.00	5.33	5.83	5.00	6.00	6.00	6.00
	Max	6.33	6.50	6.67	6.50	6.50	6.67	6.33	7.00

Letters next to the number represents statistical differences obtained by least significance difference (LSD) at $p<0.05$.

APPENDIX 2

Summary statistics of the WCR attributes found to vary with the treatments.

WCR attribute	Cultivar	Inbred				Hybrid			
	Rust control	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**
	Fruit thinning	**0%**	**0%**	**50%**	**50%**	**0%**	**0%**	**50%**	**50%**
Burnt flavor (S33)	Mean	3.50 [a]	3.50 [a]	2.00 [b]	3.00 [ab]	3.00 [ab]	2.50 [ab]	2.50 [ab]	2.50 [ab]
	Std Dev	1.29	0.58	0.00	0.82	1.41	0.58	0.58	0.58
	Min	2.00	3.00	2.00	2.00	1.00	2.00	2.00	2.00
	Max	5.00	4.00	2.00	4.00	4.00	3.00	3.00	3.00
Mouth drying astringent (S29)	Mean	2.75 [a]	2.50 [a]	2.50 [a]	2.75 [a]	1.75 [b]	2.25 [b]	2.25 [b]	2.00 [b]
	Std Dev	0.50	0.58	0.58	0.96	0.50	0.50	0.50	0.00
	Min	2.00	2.00	2.00	2.00	1.00	2.00	2.00	2.00
	Max	3.00	3.00	3.00	4.00	2.00	3.00	3.00	2.00
Overall impact flavor (S26)	Mean	9.75 [a]	9.75 [a]	8.25 [a]	9.75 [a]	8.50 [b]	8.50 [b]	8.25 [b]	8.25 [b]
	Std Dev	0.96	0.50	0.50	2.06	1.29	0.58	0.96	1.71
	Min	9.00	9.00	8.00	7.00	7.00	8.00	7.00	6.00
	Max	11.00	10.00	9.00	12.00	10.00	9.00	9.00	10.00
Musty dusty aroma (S23)	Mean	1.75 [a]	0.00 [b]	1.00 [ab]	0.00 [b]	0.50 [ab]	0.00 [b]	0.50 [ab]	0.50 [ab]
	Std Dev	1.26	0.00	2.00	0.00	1.00	0.00	1.00	1.00
	Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Max	3.00	0.00	4.00	0.00	2.00	0.00	2.00	2.00
Cardboard aroma (S19)	Mean	0.75 [ab]	1.50 [a]	0.00 [b]	0.00 [b]	0.75 [ab]	0.25 [b]	0.25 [b]	0.50 [ab]
	Std Dev	0.96	1.00	0.00	0.00	0.96	0.50	0.50	1.00
	Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Max	2.00	2.00	0.00	0.00	2.00	1.00	1.00	2.00
Stale aroma (S24)	Mean	0.00 [b]	0.00 [b]	0.50 [b]	0.00 [b]	0.00 [b]	1.25 [a]	0.00 [b]	0.25 [b]
	Std Dev	0.00	0.00	1.00	0.00	0.00	0.50	0.00	0.50
	Min	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
	Max	0.00	0.00	2.00	0.00	0.00	2.00	0.00	1.00

Letters next to the number represents statistical differences obtained by least significance difference (LSD) at $p<0.05$.

APPENDIX 3

Summary statistics of the total ion-area under curve counts of the roasted volatile compounds found to vary with the treatments.

Volatile compound	Cultivar	Inbred				Hybrid			
	Rust control	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**
	Fruit thinning	**0%**	**0%**	**50%**	**50%**	**0%**	**0%**	**50%**	**50%**
2-propionylfuran (V4)	Mean	7101 [b]	11134 [b]	86344 [ab]	76818 [ab]	3280 [b]	0 [b]	131137 [a]	0 [b]
	Std Dev	14202	22267	113269	99045	6559	0	117165	0
	Min	0	0	0	0	0	0	0	0
	Max	28403	44534	238523	188604	13118	0	277754	0
1,2,3,4-tetrahydronaphthalene-5-carboxaldehyde (V5)	Mean	34735 [b]	37416 [ab]	74089 [a]	50039 [ab]	19945 [b]	31442 [b]	37844 [ab]	43555 [ab]
	Std Dev	34215	26784	19963	36517	24405	9999	32221	20818
	Min	0	0	60566	24215	0	20269	0	20729
	Max	80203	63573	102982	91819	49779	42482	77183	63182
Ketole (V6)	Mean	36226 [b]	21713 [b]	126797 [a]	75410 [ab]	76975 [ab]	78336 [ab]	97091 [ab]	80163 [ab]
	Std Dev	45171	43426	27177	66121	65658	56374	70371	32661
	Min	0	0	93851	0	0	0	0	33510
	Max	93334	86851	150731	123456	155528	134226	154601	106087
Furfural acetone (V19)	Mean	75382 [a]	46476 [ab]	67590 [ab]	0 [ab]	0 [b]	0 [b]	68804 [ab]	0 [b]
	Std Dev	78549	53802	78049	0	0	0	58120	0
	Min	0	0	0	0	0	0	0	0
	Max	177446	97643	136005	0	0	0	137373	0
4-hydroxy-3-methylacetophenone (V21)	Mean	0 [c]	25249 [bc]	111041 [a]	92871 [ab]	20487 [bc]	24820 [bc]	62603 [abc]	24410 [bc]
	Std Dev	0	50498	76475	83888	40974	49640	78244	48821
	Min	0	0	0	0	0	0	0	0
	Max	0	100996	172324	163149	81947	99279	161878	97641
Vinylveratrole (V25)	Mean	87092 [abc]	101547 [abc]	125397 [ab]	137864 [a]	70864 [abc]	56092 [bc]	106842 [abc]	36887 [c]
	Std Dev	76930	28166	47312	16595	91004	37433	16906	35249
	Min	0	80343	60789	119290	0	0	89518	0
	Max	187414	142242	164775	151233	203302	76243	129553	78370

Appendix 3 continued…

Volatile compound	Cultivar	Inbred				Hybrid			
	Rust control	Yes	No	Yes	No	Yes	No	Yes	No
	Fruit thinning	0%	0%	50%	50%	0%	0%	50%	50%
Difurfuryl ether (V37)	Mean	284160 [ab]	337856 [ab]	458758 [a]	363051 [ab]	147193 [b]	241167 [ab]	185276 [b]	127801 [b]
	Std Dev	214127	127206	180772	156995	184833	128266	123897	86311
	Min	70312	189506	224702	217150	0	104900	79197	63196
	Max	573720	484168	662411	529179	412715	410910	359190	246919
Pentanal (V40)	Mean	0 [b]	0 [b]	1661 [ab]	1928 [ab]	0 [b]	0 [b]	3658 [a]	0 [b]
	Std Dev	0	0	3323	3339	0	0	4238	0
	Min	0	0	0	0	0	0	0	0
	Max	0	0	6645	5784	0	0	7748	0
Guaiacol (V40)	Mean	9193 [b]	0 [b]	55962 [b]	231476 [a]	23992 [b]	0 [b]	105544 [ab]	41971 [b]
	Std Dev	10885	0	65442	280948	47984	0	76381	83942
	Min	0	0	0	0	0	0	0	0
	Max	21335	0	124591	544054	95968	0	172506	167883
Phenyl acetate (V55)	Mean	15754 [c]	17590 [c]	122748 [bc]	16756 [c]	4588 [c]	782838 [a]	181198 [bc]	396201 [b]
	Std Dev	20808	35179	245497	29022	9176	133512	362397	476408
	Min	0	0	0	0	0	591240	0	0
	Max	43880	70358	490993	50267	18351	879832	724793	955178
Acetaminophen (V82)	Mean	47361 [ab]	34592 [ab]	0 [b]	48929 [ab]	22581 [b]	112114 [a]	0 [b]	26341 [ab]
	Std Dev	94723	69184	0	84748	45161	75670	0	42742
	Min	0	0	0	0	0	0	0	0
	Max	189445	138368	0	146788	90322	163388	0	89460
Coumarone (V58)	Mean	33595 [ab]	22877 [ab]	6762 [b]	25032 [ab]	7464 [b]	70003 [a]	2511 [b]	47130 [ab]
	Std Dev	39783	45754	13523	43357	14928	8138	5022	54796
	Min	0	0	0	0	0	59049	0	0
	Max	77999	91507	27046	75096	29856	77556	10043	102091
1-furfurylpyrrole (V73)	Mean	7447 [c]	16065 [bc]	55550 [a]	47153 [ab]	7733 [c]	16488 [bc]	6868 [c]	30339 [abc]
	Std Dev	14894	19100	39787	7773	15466	19058	13735	38547
	Min	0	0	0	38273	0	0	0	0
	Max	29788	37703	93575	52721	30931	34027	27470	80375

Appendix 3 continued…

Volatile compound	Cultivar	Inbred				Hybrid			
	Rust control	Yes	No	Yes	No	Yes	No	Yes	No
	Fruit thinning	0%	0%	50%	50%	0%	0%	50%	50%
2-furoylacetonitrile (V77)	Mean	0 [b]	170593 [ab]	0 [b]	145109 [ab]	0 [b]	290414 [b]	0 [b]	0 [b]
	Std Dev	0	341186	0	251336	0	347872	0	0
	Min	0	0	0	0	0	0	0	0
	Max	0	682371	0	435326	0	694150	0	0
Methanethiol (V89)	Mean	632 [b]	817 [ab]	0 [b]	1481 [ab]	945 [ab]	3293 [a]	0 [b]	1728 [ab]
	Std Dev	1263	1634	0	2565	1890	2569	0	2098
	Min	0	0	0	0	0	0	0	0
	Max	2526	3268	0	4442	3780	5496	0	4248
Pyrrole (V105)	Mean	2131 [abc]	9800 [a]	2176 [abc]	8577 [ab]	1149 [bc]	0 [c]	3436 [abc]	0 [c]
	Std Dev	4261	8384	4352	8980	2298	0	6872	0
	Min	0	0	0	0	0	0	0	0
	Max	8522	20412	8704	17912	4595	0	13743	0
2-formyl-1-methylpyrrole (V114)	Mean	99850 [a]	46128 [a]	170742 [a]	87473 [a]	10334 [ab]	0 [b]	0 [b]	0 [b]
	Std Dev	199701	92255	204941	151507	20668	0	0	0
	Min	0	0	0	0	0	0	0	0
	Max	399401	184510	410007	262418	41335	0	0	0
Camphor (V119)	Mean	21215 [ab]	9284 [b]	47366 [a]	7867 [b]	775 [b]	0 [b]	0 [b]	0 [b]
	Std Dev	42431	18567	45637	13626	1549	0	0	0
	Min	0	0	0	0	0	0	0	0
	Max	84861	37134	109665	23601	3098	0	0	0

Letters next to the number represents statistical differences obtained by least significance difference (LSD) at $p<0.05$.

APPENDIX 4

Summary statistics of the total ion-area under curve counts of the unroasted volatile compounds found to vary with the treatments.

Volatile compound	Cultivar	Inbred				Hybrid			
	Rust control	Yes	No	Yes	No	Yes	No	Yes	No
	Fruit thinning	0%	0%	50%	50%	0%	0%	50%	50%
(E)-2-Decenal	Mean	42122 [ab]	76885 [a]	0 [b]	0 [b]	0 [b]	120858 [a]	0 [b]	75241 [a]
	Std Dev	84244	88803	0	0	0	83999	0	150482
	Min	0	0	0	0	0	0	0	0
	Max	168487	156280	0	0	0	194436	0	300963
(E,E)-2,4-Decadienal	Mean	9621 [ab]	9391 [ab]	0 [b]	0 [b]	0 [b]	22572 [a]	0 [b]	10877 [ab]
	Std Dev	19241	18782	0	0	0	26472	0	21754
	Min	0	0	0	0	0	0	0	0
	Max	38482	37563	0	0	0	50812	0	43507
Nonanal	Mean	163873 [abc]	220483 [ab]	138733 [bc]	204003 [abc]	92523 [c]	257345 [a]	127888 [bc]	212347 [ab]
	Std Dev	44894	54237	102426	78563	65533	50772	97489	109949
	Min	112991	167441	0	138721	0	200731	0	152681
	Max	220753	293383	224317	306584	140420	312598	220404	376957
Ethanol	Mean	321173 [ab]	410526 [ab]	502263 [ab]	371916 [ab]	468675 [ab]	107144 [b]	639503 [a]	487757 [ab]
	Std Dev	209217	300996	255864	302342	368315	156293	302300	361520
	Min	136426	129992	188446	0	191845	0	447039	138445
	Max	620208	799285	732228	732228	1010969	331257	1089430	989363
Dimethyl sulfide	Mean	0 [b]	0 [b]	21841 [b]	0 [b]	0 [b]	113739 [a]	0 [b]	0 [b]
	Std Dev	0	0	43681	0	0	227477	0	0
	Min	0	0	0	0	0	0	0	0
	Max	0	0	87362	0	0	454954	0	0

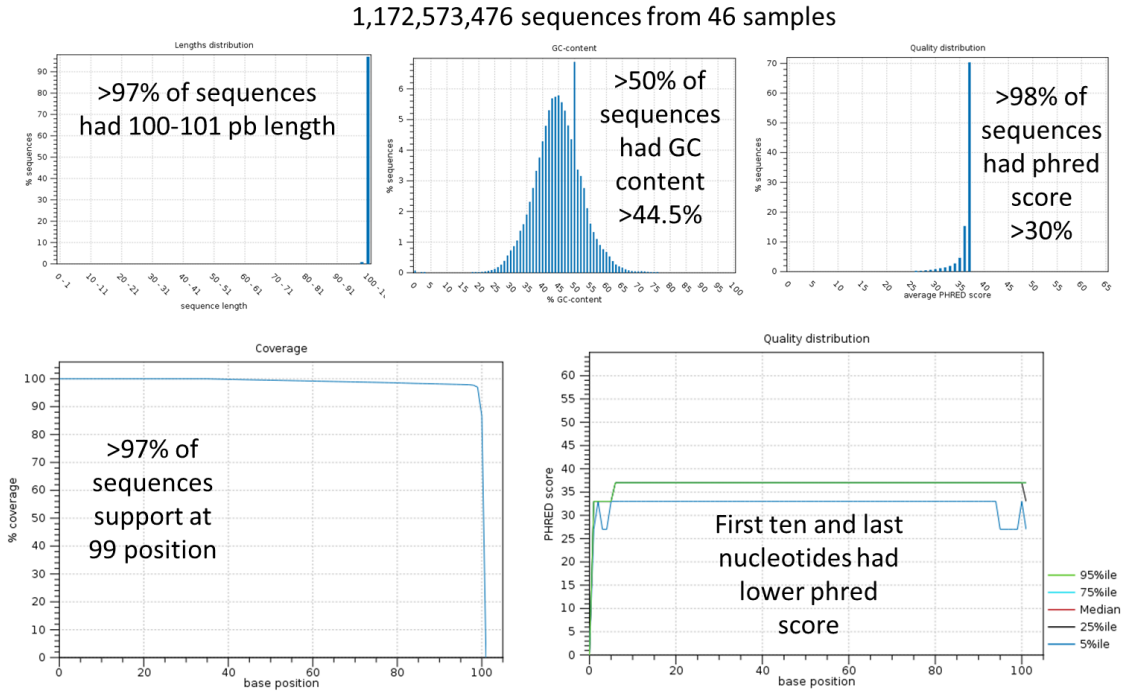
Appendix 4 continued…

Volatile compound	Cultivar	Inbred				Hybrid			
	Rust control	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**
	Fruit thinning	**0%**	**0%**	**50%**	**50%**	**0%**	**0%**	**50%**	**50%**
Acetic acid	Mean	0 [b]	0 [b]	0 [b]	0 [b]	30830 [a]	0 [b]	11277 [a]	53785 [a]
	Std Dev	0	0	0	0	61661	0	22555	71025
	Min	0	0	0	0	0	0	0	0
	Max	0	0	0	0	123321	0	45109	149775
dl-Limonene	Mean	0 [b]	0 [b]	0 [b]	0 [b]	36361 [a]	0 [b]	30320 [a]	19893 [a]
	Std Dev	0	0	0	0	72722	0	60640	39786
	Min	0	0	0	0	0	0	0	0
	Max	0	0	0	0	145443	0	121279	79571

Letters next to the number represents statistical differences obtained by least significant difference (LSD) at $p<0.05$.

APPENDIX 5

Summary of the quality report of the fruit sequences after trimming.



APPENDIX 6

Quality report of the fruit fragments and mapping of the samples according to the treatments. The total, standard deviation (sd), and percent of each parameter are described for each treatment and maturity stage. Treatment number corresponds to Table 3.

	Parameter	Treatment 1			Treatment 2			Treatment 3			Treatment 4			Treatment 5			Treatment 6			Treatment 7			Treatment 8			Average immature		
		Total	sd	%	Total	sd	%	Total	sd	%	Total	sd	%	Total	sd	%	Total	sd	%	Total	sd	%	Total	sd	%	Total	sd	%
Immature	Counted fragments	2.6E+07	2.3E+06	98	2.0E+07	1.7E+06	96	2.4E+07	3.9E+06	97	2.4E+07	1.8E+06	98	2.3E+07	1.5E+06	97	2.7E+07	5.0E+06	97	2.1E+07	1.9E+06	96	2.0E+07	1.2E+06	93	2.3E+07	2.5E+06	96
	unique fragments	2.6E+07	2.3E+06	95	2.0E+07	1.6E+06	94	2.3E+07	3.8E+06	95	2.3E+07	1.7E+06	95	2.2E+07	1.5E+06	95	2.7E+07	4.9E+06	94	2.0E+07	1.9E+06	94	2.0E+07	1.3E+06	90	2.3E+07	2.4E+06	94
	non-specifically	7.7E+05	1.5E+05	3	5.6E+05	7.9E+04	3	6.5E+05	8.2E+04	3	6.2E+05	6.8E+04	3	5.7E+05	1.6E+04	2	7.8E+05	8.7E+04	3	5.5E+05	3.8E+04	3	5.4E+05	1.8E+04	2	6.3E+05	9.1E+04	3
	Uncounted fragments	6.2E+05	1.6E+05	2	7.9E+05	3.3E+05	4	6.3E+05	3.0E+04	3	5.5E+05	1.4E+05	2	6.4E+05	2.0E+05	3	8.8E+05	3.5E+05	3	7.9E+05	3.3E+05	4	1.6E+06	9.2E+05	7	8.2E+05	3.2E+05	4
	Total fragments	2.7E+07	2.2E+06	100	2.1E+07	1.7E+06	100	2.4E+07	3.9E+06	100	2.5E+07	1.7E+06	100	2.3E+07	1.4E+06	100	2.8E+07	4.8E+06	100	2.1E+07	2.0E+06	100	2.2E+07	3.2E+05	100	2.4E+07	2.4E+06	100
		Mapped	sd	%	Mapped	sd	%	Mapped	sd	%	Mapped		%	Mapped	sd	%	Mapped	sd	%	Mapped		%	Mapped	sd	%	Mapped	sd	%
	Exon	1.6E+07	7.1E+05	59	1.2E+07	1.1E+06	58	1.3E+07	2.1E+06	57	1.4E+07	9.6E+05	57	1.3E+07	1.1E+06	56	1.5E+07	2.8E+06	55	1.2E+07	1.3E+06	58	1.2E+07	7.1E+05	58	1.3E+07	1.4E+06	57
	Exon-exon	6.6E+06	6.3E+05	25	5.0E+06	3.6E+05	25	6.0E+06	1.0E+06	25	6.0E+06	4.2E+05	25	6.0E+06	5.7E+05	26	6.8E+06	1.3E+06	25	5.1E+06	6.7E+05	25	4.8E+06	4.1E+05	24	5.8E+06	6.8E+05	25
	Total exon	2.2E+07	1.3E+06	84	1.7E+07	1.5E+06	83	1.9E+07	3.1E+06	82	2.0E+07	1.3E+06	83	1.9E+07	1.6E+06	83	2.2E+07	4.1E+06	80	1.7E+07	2.0E+06	82	1.7E+07	1.1E+06	81	1.9E+07	2.0E+06	82
	Intron	1.4E+06	5.5E+05	5	1.3E+06	5.2E+04	6	1.4E+06	2.5E+05	6	1.4E+06	1.4E+05	6	1.4E+06	3.2E+04	6	1.7E+06	2.6E+05	6	1.1E+06	1.8E+05	5	1.1E+06	1.9E+05	5	1.4E+06	1.8E+05	6
	Total gene	2.4E+07	1.9E+06	89	1.8E+07	1.5E+06	89	2.1E+07	3.3E+06	88	2.1E+07	1.5E+06	88	2.0E+07	1.7E+06	89	2.4E+07	4.3E+06	86	1.8E+07	2.1E+06	88	1.8E+07	1.3E+06	87	2.0E+07	2.2E+06	88
	Intergenic	2.8E+06	5.7E+05	11	2.2E+06	2.3E+05	11	2.8E+06	5.9E+05	12	2.8E+06	4.0E+05	12	2.4E+06	2.5E+05	11	3.8E+06	6.9E+05	14	2.5E+06	5.7E+05	12	2.7E+06	2.8E+05	13	2.8E+06	4.5E+05	12
	Total	2.6E+07	2.3E+06	100	2.0E+07	1.7E+06	100	2.4E+07	3.9E+06	100	2.4E+07	1.8E+06	100	2.3E+07	1.5E+06	100	2.7E+07	5.0E+06	100	2.1E+07	1.9E+06	100	2.0E+07	1.2E+06	100	2.3E+07	2.5E+06	100

	Parameter	Treatment 1			Treatment 2			Treatment 3			Treatment 4			Treatment 5			Treatment 6			Treatment 7			Treatment 8			Average mature		
		Total	sd	%	Total	sd	%	Total	sd	%	Total	sd	%	Total	sd	%	Total	sd	%	Total	sd	%	Total	sd	%	Total	sd	%
Mature	Counted fragments	2.4E+07	2.0E+06	98	2.3E+07	1.5E+06	94	3.4E+07	1.1E+07	93	2.7E+07	3.5E+06	96	2.5E+07	1.1E+07	96	2.7E+07	8.1E+06	98	2.3E+07	5.3E+06	96	2.4E+07	3.2E+06	97	2.6E+07	3.5E+06	96
	unique fragments	2.4E+07	2.0E+06	96	2.2E+07	1.4E+06	92	3.4E+07	1.1E+07	90	2.7E+07	3.6E+06	94	2.5E+07	1.0E+07	93	2.6E+07	7.9E+06	95	2.2E+07	5.2E+06	93	2.3E+07	3.1E+06	94	2.5E+07	3.4E+06	93
	non-specifically	5.6E+05	4.1E+04	2	6.1E+05	9.0E+04	2	8.5E+05	2.1E+05	2	7.8E+05	8.9E+04	3	6.2E+05	2.4E+05	2	6.5E+05	2.1E+05	2	6.3E+05	1.4E+05	3	6.3E+05	1.0E+05	3	6.6E+05	9.1E+04	2
	Uncounted fragments	4.5E+05	8.8E+04	2	1.4E+06	7.8E+05	6	2.4E+06	1.4E+06	7	9.6E+05	3.4E+05	4	1.3E+06	8.5E+05	4	6.2E+05	2.6E+05	2	1.0E+06	3.9E+05	4	8.0E+05	3.9E+05	3	1.1E+06	5.8E+05	4
	Total fragments	2.5E+07	1.9E+06	100	2.4E+07	1.7E+06	100	3.7E+07	1.1E+07	100	2.8E+07	3.2E+06	100	2.7E+07	1.1E+07	100	2.7E+07	8.4E+06	100	2.4E+07	5.4E+06	100	2.5E+07	3.4E+06	100	2.7E+07	4.0E+06	100
		Mapped	sd	%	Mapped	sd	%	Mapped	sd	%	Mapped		%	Mapped	sd	%	Mapped	sd	%	Mapped	sd	%	Mapped	sd	%	Mapped	sd	%
	Exon	1.4E+07	1.3E+06	59	1.4E+07	1.1E+06	59	2.0E+07	6.1E+06	59	1.6E+07	2.5E+06	59	1.5E+07	6.2E+06	58	1.5E+07	4.7E+06	58	1.4E+07	3.8E+06	58	1.4E+07	1.8E+06	58	1.5E+07	2.0E+06	58
	Exon-exon	6.2E+06	4.8E+05	26	5.3E+06	5.8E+05	23	8.3E+06	3.6E+06	23	6.0E+06	1.7E+06	21	6.6E+06	2.9E+06	26	6.7E+06	2.1E+06	25	5.7E+06	1.5E+06	25	6.0E+06	7.5E+05	25	6.3E+06	8.6E+05	24
	Total exon	2.1E+07	1.8E+06	84	1.9E+07	1.1E+06	82	2.8E+07	9.7E+06	82	2.2E+07	4.2E+06	80	2.1E+07	9.1E+06	83	2.2E+07	6.8E+06	83	1.9E+07	5.3E+06	83	2.0E+07	2.6E+06	83	2.2E+07	2.9E+06	82
	Intron	1.6E+06	9.5E+04	7	1.2E+06	1.7E+05	5	2.1E+06	9.3E+05	6	1.4E+06	4.1E+05	5	1.6E+06	7.3E+05	6	1.8E+06	5.9E+05	7	1.1E+06	1.8E+05	5	1.5E+06	1.4E+05	6	1.5E+06	2.9E+05	6
	Total gene	2.2E+07	1.8E+06	91	2.0E+07	1.1E+06	87	3.1E+07	1.1E+07	87	2.3E+07	4.6E+06	85	2.3E+07	9.8E+06	90	2.4E+07	7.4E+06	90	2.0E+07	5.2E+06	88	2.1E+07	2.7E+06	89	2.3E+07	3.1E+06	88
	Intergenic	2.2E+06	1.5E+05	9	3.0E+06	1.1E+06	13	4.0E+06	2.1E+05	13	3.9E+06	1.1E+06	15	2.5E+06	7.5E+05	10	2.7E+06	7.7E+05	10	2.6E+06	5.1E+05	12	2.7E+06	5.9E+05	11	3.0E+06	6.0E+05	12
	Total	2.4E+07	2.0E+06	100	2.3E+07	1.5E+06	100	3.4E+07	1.1E+07	100	2.7E+07	3.5E+06	100	2.5E+07	1.1E+07	100	2.7E+07	8.1E+06	100	2.3E+07	5.3E+06	100	2.4E+07	3.2E+06	100	2.6E+07	3.5E+06	100

APPENDIX 7

Sequencing depth distribution for the immature (A) and mature (B) bean samples.

A.

Tissue	Immature										
Cultivar	Inbred				Hybrid				Inbred	Hybrid	Overall
Treatment	R	C	R+T	T	R	C	R+T	T			
>0-0.5X	16,903	18,257	17,836	17,721	18,426	17,750	18,919	19,091	19,004	19,875	20,043
0.5-5X	4,471	3,505	3,764	3,787	3,607	4,355	3,338	3,054	3,909	3,600	3,746
5X-50X	354	271	285	308	246	332	237	241	304	258	286
50X-500X	19	13	15	18	13	19	13	15	16	15	15
>500X	1	1	1	1	1	1	1	1	1	1	1
Total higher than 0.5X	4,845	3,790	4,065	4,114	3,867	4,707	3,589	3,311	4,230	3,874	4,048
Percentage higher than 0.5X	22.3%	17.2%	18.6%	18.8%	17.3%	21.0%	15.9%	14.8%	18.2%	16.3%	16.8%
Total expressed genes	21,748	22,047	21,901	21,835	22,293	22,457	22,508	22,402	23,234	23,749	24,091

B.

Tissue	Mature										
Cultivar	Inbred				Hybrid				Inbred	Hybrid	Overall
Treatment	R	C	R+T	T	R	C	R+T	T			
>0-0.5X	17,376	18,107	16,557	17,422	17,486	17,634	17,928	17,936	18,770	19,177	17,839
0.5-5X	4,158	3,712	5,155	4,259	4,369	4,495	3,358	3,942	4,355	4,051	4,185
5X-50X	297	290	470	361	332	325	215	309	357	305	326
50X-500X	16	15	29	18	17	15	10	17	18	15	17
>500X	1	1	2	1	1	1	1	1	1	1	1
Total higher than 0.5X	4,472	4,018	5,656	4,639	4,719	4,836	3,584	4,269	4,731	4,372	4,530
Percentage higher than 0.5X	20.5%	18.2%	25.5%	21.0%	21.3%	21.5%	16.7%	19.2%	20.1%	18.6%	20.2%
Total expressed genes	21,848	22,125	22,213	22,061	22,205	22,470	21,512	22,205	23,501	23,549	22,369

APPENDIX 8

Total 471 differentially expressed genes (DEGs) found when compared immature vs mature bean stages. Table where reported in the dissertation is specified in the first column. Gene ID and annotation are displayed according to the reference genome [134]. Average sequencing depth was made by averaging all samples and treatments as described by Dugas, et al. [136]. The GO term number was presented for known genes using AgriGO 2.0 [188]. Fold change (FC) higher than 2.0 represents up-regulated expression (increased under mature stage), while lower than 2.0 represents down-regulated expression (increased under immature stage). Statistical significance using Bonferroni and FDR are shown.

Reported in:	**Gene ID [134]**	**Annotation description**	**Average sequencing depth**	**GO term**	**Fold change**	**Bonferroni**	**FDR *p*-value correction**
	Cc00_g00230	Putative uncharacterized protein	0.010	GO:0003677	-4.073	0.000	0.000
Table 8	Cc00_g00780	Putative 14 kDa proline-rich protein DC2.15	0.172	GO:0006869	524.085	0.000	0.000
Table 8	Cc00_g00800	Putative 14 kDa proline-rich protein DC2.15	0.169	GO:0006869	436.226	0.000	0.000
	Cc00_g00970	Germin-like protein subfamily 1 member 7	0.026	GO:0048046	-3.392	0.000	0.000
	Cc00_g01250	Putative Cytochrome P450 71A1	0.003	GO:0004497	-3.207	0.001	0.000
	Cc00_g02000	Mitochondrial transcription termination factor family protein	0.009	N/A	-4.095	0.000	0.000
	Cc00_g04040	60S ribosomal protein L27a-2	0.052	N/A	63.691	0.001	0.000
Table 8	Cc00_g04320	Putative Disease resistance-responsive (dirigent-like protein) family protein	0.113	N/A	344.208	0.000	0.000
	Cc00_g05170	Putative Cytochrome P450 86B1	0.009	GO:0004497	-2.885	0.001	0.000
	Cc00_g05810	Hypothetical protein	0.105	N/A	-4.894	0.000	0.000
	Cc00_g08270	Isoflavone reductase homolog A622	0.030	GO:0003824	-2.449	0.003	0.000
	Cc00_g09660	Beta-glucosidase 12	0.034	GO:0003824	-8.149	0.000	0.000
	Cc00_g10320	Putative Protein ASPARTIC PROTEASE IN GUARD CELL 1	0.019	GO:0006508	-5.313	0.000	0.000
	Cc00_g10900	Oligopeptide transporter 4	0.021	GO:0055085	-3.501	0.000	0.000
	Cc00_g10910	Oligopeptide transporter 4	0.087	GO:0055085	-2.879	0.000	0.000
	Cc00_g11850	Major facilitator superfamily protein	0.010	GO:0016020	-4.545	0.000	0.000
	Cc00_g12470	Putative Probable flavonol synthase 5	0.018	GO:0016491	-3.071	0.002	0.000
	Cc00_g12540	Hypothetical protein	0.075	N/A	2.743	0.004	0.000
Table 8	Cc00_g12970	8-hydroxyquercetin 8-O-methyltransferase	0.014	GO:0046983	4.736	0.000	0.000

Appendix 8 continued…

Reported in:	Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
	Cc00_g12980	8-hydroxyquercetin 8-O-methyltransferase	0.984	GO:0046983	2.663	0.003	0.000
	Cc00_g13640	Momilactone A synthase	0.021	GO:0055114	200.737	0.002	0.000
	Cc00_g13740	Probable pectinesterase/pectinesterase inhibitor 12	0.017	GO:0005618	-3.290	0.001	0.000
	Cc00_g13750	Probable pectinesterase/pectinesterase inhibitor 32	0.001	GO:0005618	-9.054	0.000	0.000
	Cc00_g13760	Pectinesterase/pectinesterase inhibitor PPE8B	0.113	GO:0005618	-2.937	0.000	0.000
	Cc00_g14010	3-ketoacyl-CoA synthase 21	0.003	GO:0006633	-20.997	0.000	0.000
	Cc00_g14150	Protein COBRA	0.352	N/A	-2.142	0.000	0.000
	Cc00_g16430	8-hydroxyquercetin 8-O-methyltransferase	0.420	GO:0008171	3.408	0.000	0.000
	Cc00_g17010	Putative Probable glucan 1,3-beta-glucosidase A	0.017	GO:0003824	-5.637	0.000	0.000
	Cc00_g17860	Probable polygalacturonase	0.007	GO:0005975	-3.235	0.001	0.000
	Cc00_g18340	unknown protein; FUNCTIONS IN	0.182	N/A	-2.018	0.008	0.000
	Cc00_g18470	G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	0.002	GO:0006468	-3.642	0.000	0.000
	Cc00_g19050	Beta-glucosidase 11	0.007	GO:0003824	-2.599	0.000	0.000
	Cc00_g19230	Putative Early nodulin-like protein 2	0.074	GO:0005507	-2.014	0.001	0.000
	Cc00_g20120	AAA-ATPase 1	0.085	GO:0017111	-2.250	0.000	0.000
	Cc00_g21130	Putative Thiol protease SEN102	0.023	GO:0006508	-126.136	0.000	0.000
	Cc00_g21310	Putative Probable leucine-rich repeat receptor-like protein kinase At1g35710	0.001	GO:0006468	-7.024	0.001	0.000
	Cc00_g21470	Aromatic-L-amino-acid decarboxylase	10.590	GO:0003824	-3.009	0.000	0.000
	Cc00_g22300	Dammarenediol II synthase	0.012	N/A	110.648	0.002	0.000
Table 8	Cc00_g23280	Aluminum-activated malate transporter 10	0.003	GO:0010044	-6.934	0.000	0.000
	Cc00_g23580	Isoflavone reductase homolog A622	0.072	GO:0003824	-2.127	0.001	0.000
	Cc00_g23650	Putative Flavonoid 3'-monooxygenase	0.012	GO:0004497	-6.977	0.003	0.000
	Cc00_g24100	GDSL esterase/lipase At1g28580	0.742	GO:0006629	-2.963	0.000	0.000
	Cc00_g24680	Bidirectional sugar transporter SWEET10	0.005	GO:0016021	-3.097	0.007	0.000
	Cc00_g24740	unknown protein; FUNCTIONS IN	0.003	N/A	-4.779	0.001	0.000
	Cc00_g25070	Endoglucanase 12	0.151	GO:0003824	-3.405	0.000	0.000
	Cc00_g25270	Putative Isoflavone 2'-hydroxylase	0.029	GO:0004497	125.282	0.005	0.000

Appendix 8 continued…

Reported in:	Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
	Cc00_g25490	Putative 8-hydroxyquercetin 8-O-methyltransferase	0.783	GO:0046983	2.822	0.004	0.000
	Cc00_g25920	Phosphate-responsive 1 family protein	0.195	N/A	-2.643	0.001	0.000
	Cc00_g25930	Phosphate-responsive 1 family protein	0.022	N/A	-4.760	0.000	0.000
	Cc00_g26170	Glutathione S-transferase F12	0.027	N/A	3.068	0.000	0.000
	Cc00_g26210	Hypothetical protein	0.037	N/A	-10.814	0.000	0.000
	Cc00_g26450	alpha/beta-Hydrolases superfamily protein	0.002	GO:0004806	-10.119	0.000	0.000
	Cc00_g27010	UDP-glucuronate 4-epimerase 1	0.044	GO:0050662	-2.522	0.000	0.000
	Cc00_g27020	Salicylate O-methyltransferase	0.056	GO:0008168	-3.621	0.007	0.000
	Cc00_g27510	Putative unknown seed protein like 1	0.030	N/A	-2.675	0.002	0.000
	Cc00_g28090	Ribulose bisphosphate carboxylase small chain A, chloroplastic	0.191	GO:0015977	-22.125	0.009	0.000
	Cc00_g28850	Caffeic acid 3-O-methyltransferase	0.006	GO:0046983	-4.524	0.003	0.000
	Cc00_g29100	Putative 8-hydroxyquercetin 8-O-methyltransferase	0.007	GO:0046983	7.215	0.000	0.000
	Cc00_g30820	60S ribosomal protein L39-1	0.104	GO:0003735	2.431	0.000	0.000
	Cc00_g32530	Putative O-methyltransferase family protein	0.006	GO:0046983	7.797	0.000	0.000
	Cc00_g35890	Polyphenol oxidase I, chloroplastic	0.255	GO:0008152	-6.753	0.000	0.000
	Cc01_g00720	3,7-dimethylxanthine N-methyltransferase	0.323	GO:0008168	-8.019	0.000	0.000
	Cc01_g01210	Putative Disease resistance response protein 206	0.040	N/A	-3.224	0.000	0.000
	Cc01_g01530	Family of unknown function (DUF716)	0.091	N/A	-3.584	0.000	0.000
	Cc01_g01570	Family of unknown function (DUF716)	0.005	N/A	-35.430	0.000	0.000
	Cc01_g02550	Putative F-box protein At5g07610	0.009	N/A	3.116	0.005	0.000
	Cc01_g03630	Putative uncharacterized protein	0.006	GO:0016567	-2.018	0.005	0.000
	Cc01_g03860	Putative Flavonol synthase/flavanone 3-hydroxylase	0.006	GO:0016491	-4.665	0.000	0.000
	Cc01_g03970	Isoflavone reductase homolog P3	0.154	GO:0003824	-2.477	0.001	0.000
	Cc01_g04060	Linoleate 13S-lipoxygenase 2-1, chloroplastic	0.171	GO:0005506	-16.998	0.000	0.000
	Cc01_g04600	Putative Auxin-induced protein 5NG4	0.012	GO:0016020	-6.726	0.000	0.000
	Cc01_g05870	Xyloglucan glycosyltransferase 4	0.014	N/A	-2.690	0.000	0.000
	Cc01_g06970	Putative Hyoscyamine 6-dioxygenase	0.025	GO:0016491	-205.066	0.000	0.000

Appendix 8 continued…

Reported in:	Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
	Cc01_g07410	Alcohol dehydrogenase-like 7	1.403	GO:0008270	-2.051	0.000	0.000
	Cc01_g07420	Alcohol dehydrogenase-like 7	3.732	GO:0008270	-2.092	0.000	0.000
	Cc01_g08380	Putative Lysosomal beta glucosidase	0.131	GO:0005975	-2.279	0.000	0.000
	Cc01_g08500	Aluminum-activated malate transporter 2	0.019	GO:0010044	-3.505	0.000	0.000
	Cc01_g08680	Momilactone A synthase	0.337	GO:0055114	-2.453	0.000	0.000
	Cc01_g08940	Germin-like protein 9-3	0.012	GO:0048046	59.387	0.002	0.000
	Cc01_g09500	Putative Cytochrome P450 94A1	0.003	GO:0004497	-3.890	0.005	0.000
	Cc01_g09780	Arogenate dehydratase/prephenate dehydratase 6, chloroplastic	0.040	GO:0004664	-2.186	0.000	0.000
	Cc01_g09870	Putative Hyoscyamine 6-dioxygenase	0.118	GO:0016491	-18.278	0.000	0.000
	Cc01_g09890	Glucomannan 4-beta-mannosyltransferase 2	0.132	N/A	-2.763	0.001	0.000
	Cc01_g10180	Hypothetical protein	0.004	N/A	-5.530	0.001	0.000
	Cc01_g10430	Serine carboxypeptidase-like 48	0.005	GO:0006508	-5.231	0.007	0.000
	Cc01_g10770	Putative Cysteine proteinases superfamily protein	0.025	N/A	-2.331	0.000	0.000
	Cc01_g10810	Putative Receptor-like protein 12	0.007	GO:0005515	-2.961	0.000	0.000
	Cc01_g10940	Putative cyclic nucleotide-gated ion channel 15	0.002	GO:0006813	-3.460	0.000	0.000
	Cc01_g11870	Mitogen-activated protein kinase homolog NTF6	0.009	GO:0004674	-2.839	0.000	0.000
	Cc01_g12010	Histone H4	0.438	GO:0005634	-3.315	0.007	0.000
	Cc01_g12260	Aquaporin TIP1-1	0.056	GO:0005215	-2.832	0.001	0.000
	Cc01_g12350	Putative F-box/kelch-repeat protein At1g80440	0.126	N/A	-2.051	0.000	0.000
	Cc01_g12380	Probable histone H2B.1	0.258	GO:0005634	-3.481	0.000	0.000
	Cc01_g13840	Peptide transporter PTR5	0.008	GO:0006857	-4.022	0.000	0.000
	Cc01_g14520	Putative uncharacterized protein	0.002	GO:0008270	-5.560	0.001	0.000
	Cc01_g15540	Putative cytochrome P450, family 714, subfamily A, polypeptide 1	0.010	GO:0004497	-5.329	0.000	0.000
	Cc01_g15550	Putative Cytochrome P450 734A6	0.011	GO:0004497	-4.470	0.000	0.000
	Cc01_g16450	Putative Probable glucan endo-1,3-beta-glucosidase A6	0.013	GO:0003824	-2.436	0.001	0.000

Appendix 8 continued…

Reported in:	Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
	Cc01_g16870	Xyloglucan endotransglucosylase/hydrolase protein 2	0.135	GO:0005618	-4.487	0.000	0.000
	Cc01_g17180	Putative CLAVATA3/ESR (CLE)-related protein TDIF	0.074	N/A	-2.614	0.000	0.000
	Cc01_g18220	NAC domain protein, IPR003441	0.034	GO:0045449	-2.229	0.002	0.000
	Cc01_g18410	Putative Hyoscyamine 6-dioxygenase	0.017	GO:0016491	-3.061	0.003	0.000
	Cc01_g18420	Putative Hyoscyamine 6-dioxygenase	0.041	GO:0016491	-2.559	0.000	0.000
	Cc01_g18710	AP2/ERF domain-containing transcription factor	0.006	GO:0003700	-7.783	0.000	0.000
	Cc01_g18830	Phospholipase A1-Igamma2, chloroplastic	0.038	GO:0004806	-4.679	0.000	0.000
	Cc01_g19130	Sulfate transporter 3.1	0.007	GO:0005215	-2.451	0.007	0.000
	Cc01_g19640	Putative Basic 7S globulin	1.912	GO:0006508	-2.267	0.000	0.000
	Cc01_g19650	Eukaryotic aspartyl protease family protein	0.896	GO:0006508	-2.315	0.000	0.000
	Cc01_g19890	DNAse I-like superfamily protein	0.002	N/A	-3.055	0.003	0.000
	Cc01_g20330	Bidirectional sugar transporter SWEET14	0.154	GO:0016021	-9.782	0.000	0.000
	Cc02_g02070	Acidic endochitinase SE2	2.456	GO:0003824	-2.685	0.000	0.000
	Cc02_g03510	Putative peptidoglycan-binding LysM domain-containing protein	0.016	GO:0016998	2.422	0.002	0.000
	Cc02_g03540	Protein HOTHEAD	0.027	GO:0006066	-2.358	0.000	0.000
	Cc02_g03710	Putative 3-ketoacyl-CoA synthase 5	0.009	GO:0006633	-10.463	0.000	0.000
	Cc02_g04520	Putative leucine-rich repeat receptor-like protein kinase At2g19210	0.009	GO:0004674	-2.210	0.000	0.000
	Cc02_g05640	Peroxidase 4	0.019	GO:0055114	-10.501	0.000	0.000
	Cc02_g05900	unknown protein; BEST Arabidopsis thaliana protein match is	0.018	N/A	-4.812	0.000	0.000
	Cc02_g06760	Putative uncharacterized protein	0.043	N/A	-2.071	0.003	0.000
	Cc02_g07170	Putative UDP-glycosyltransferase 85A2	0.194	GO:0008152	-2.187	0.000	0.000
	Cc02_g08440	Putative Protein of unknown function (DUF674)	0.016	N/A	-2.490	0.000	0.000
	Cc02_g08610	ABC transporter G family member 1	0.037	GO:0016020	-2.121	0.000	0.000
	Cc02_g08880	Protein of unknown function (DUF640)	0.129	N/A	-2.326	0.008	0.000
	Cc02_g09000	Putative Subtilisin-like protease	0.008	GO:0004252	-49.964	0.000	0.000

Appendix 8 continued…

Reported in:	Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
	Cc02_g09150	GDSL esterase/lipase At5g33370	0.007	GO:0006629	-7.925	0.007	0.000
	Cc02_g09430	Inositol oxygenase 1	0.055	GO:0005737	-9.494	0.000	0.000
	Cc02_g10760	Major facilitator superfamily protein	0.365	GO:0055085	-2.401	0.004	0.000
	Cc02_g11380	Putative Predicted protein	0.025	N/A	-3.063	0.000	0.000
	Cc02_g11660	Probable xyloglucan endotransglucosylase/hydrolase protein 8	0.073	GO:0005618	-3.551	0.000	0.000
	Cc02_g11870	(3S,6E)-nerolidol synthase 1, chloroplastic	0.387	GO:0016829	2.412	0.003	0.000
	Cc02_g12170	Uncharacterized protein	0.463	GO:0008270	-2.495	0.000	0.000
	Cc02_g12240	Major facilitator superfamily protein	0.011	GO:0016020	-4.740	0.000	0.000
	Cc02_g12250	Major facilitator superfamily protein	0.038	GO:0016020	-3.112	0.001	0.000
	Cc02_g12260	Major facilitator superfamily protein	0.003	GO:0016020	-4.829	0.000	0.000
	Cc02_g12330	Putative Transcription factor MYB39	0.057	GO:0045449	-4.814	0.000	0.000
	Cc02_g12630	Putative Probable NADH dehydrogenase	0.045	GO:0050660	-2.847	0.000	0.000
	Cc02_g12790	Putative S-linalool synthase	0.006	GO:0016829	2.435	0.000	0.000
	Cc02_g14240	Pathogenesis-related genes transcriptional activator PTI5	0.146	GO:0003700	-2.735	0.000	0.000
	Cc02_g14250	Putative Ethylene-responsive transcription factor ERF091	0.006	GO:0003700	-2.922	0.001	0.000
	Cc02_g14610	Phospholipase A1-IIgamma	0.010	GO:0004806	-13.657	0.002	0.000
	Cc02_g14650	TRICHOME BIREFRINGENCE-LIKE 38	39.965	N/A	-2.017	0.000	0.000
	Cc02_g14960	Dof zinc finger protein DOF1.5	0.069	GO:0045449	-4.176	0.000	0.000
	Cc02_g16170	GDSL esterase/lipase At1g29670	0.005	GO:0006629	-11.205	0.002	0.000
	Cc02_g16680	Putative HVA22-like protein g	0.014	N/A	-3.552	0.000	0.000
	Cc02_g17170	Beta-galactosidase 3	0.012	GO:0003824	-2.596	0.000	0.000
	Cc02_g17620	Glucan endo-1,3-beta-glucosidase 7	0.028	GO:0003824	-2.710	0.000	0.000
	Cc02_g18650	Putative uncharacterized protein	0.060	GO:0007050	2.451	0.000	0.000
	Cc02_g20390	Receptor-like protein kinase HAIKU2	0.117	GO:0004674	-2.085	0.000	0.000
	Cc02_g21300	UDP-glycosyltransferase 85A2	0.121	GO:0008152	-3.962	0.000	0.000
	Cc02_g21680	Cationic peroxidase 2	0.001	GO:0055114	-5.159	0.004	0.000
	Cc02_g21830	Beta-xylosidase/alpha-L-arabinofuranosidase 2	1.010	GO:0005975	-3.101	0.000	0.000
	Cc02_g22030	Putative Auxin-induced protein 5NG4	0.042	GO:0016020	-4.336	0.000	0.000

Appendix 8 continued ..

Reported in:	Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
	Cc02_g22890	COBRA-like protein 6	0.051	N/A	-2.312	0.000	0.000
	Cc02_g24030	UDP-glucuronate 4-epimerase 3	0.109	GO:0050662	-2.563	0.000	0.000
	Cc02_g24270	Hypothetical protein	10.052	N/A	-16.650	0.000	0.000
	Cc02_g24440	SPX domain-containing protein 3	0.114	N/A	-3.685	0.000	0.000
	Cc02_g24560	Putative Metalloendoproteinase 1	0.027	GO:0004222	4.845	0.000	0.000
	Cc02_g25050	Hypothetical protein	3.151	N/A	-4.522	0.000	0.000
	Cc02_g25800	Putative Protein WAX2	0.015	GO:0006633	-2.522	0.000	0.000
	Cc02_g26520	Putative Protein PHLOEM PROTEIN 2-LIKE A1	0.142	N/A	-2.646	0.000	0.000
	Cc02_g26760	Putative Domain of unknown function (DUF966)	0.012	N/A	-3.025	0.000	0.000
	Cc02_g26820	Putative Major facilitator superfamily domain-containing protein 12	0.030	N/A	-2.252	0.000	0.000
	Cc02_g27410	Histone H3.2	0.173	GO:0006334	-2.729	0.000	0.000
	Cc02_g28090	Putative Polygalacturonase QRT2	0.009	GO:0005975	-4.226	0.000	0.000
	Cc02_g28890	Serine carboxypeptidase-like 40	0.059	GO:0006508	-2.858	0.001	0.000
	Cc02_g30100	Cytokinin dehydrogenase 9	0.052	GO:0050660	-17.011	0.000	0.000
	Cc02_g32290	Putative unknown protein; Has 23 Blast hits to 23 proteins in 9 species	0.329	N/A	-2.629	0.000	0.000
Table 8	Cc02_g33130	Respiratory burst oxidase homolog protein B	0.009	GO:0050660	-8.430	0.000	0.000
	Cc02_g33210	Luminal-binding protein 5	0.006	GO:0005524	-2.747	0.003	0.000
	Cc02_g33320	Probable linoleate 9S-lipoxygenase 5	0.371	GO:0005506	-3.131	0.000	0.000
	Cc02_g33670	rotamase CYP 1	0.018	GO:0006457	-16.830	0.004	0.000
	Cc02_g33800	Probable linoleate 9S-lipoxygenase 5	0.071	GO:0005506	-5.944	0.000	0.000
	Cc02_g33810	Hypothetical protein	0.028	GO:0046872	-4.535	0.001	0.000
	Cc02_g34690	Putative uncharacterized protein	0.052	N/A	-3.236	0.000	0.000
	Cc02_g34950	Putative Predicted protein	0.013	N/A	3.437	0.000	0.000
	Cc02_g35270	Protein of unknown function, DUF538	1.855	N/A	-2.167	0.000	0.000
	Cc02_g35450	GDSL esterase/lipase At1g09390	0.015	GO:0006629	-2.228	0.002	0.000
	Cc02_g35720	Pleiotropic drug resistance protein 3	0.021	GO:0016020	-3.022	0.000	0.000
	Cc02_g36510	Fasciclin-like arabinogalactan protein 8	0.051	N/A	-3.613	0.000	0.000
	Cc02_g36980	MADS-box transcription factor 6	0.031	GO:0003700	-2.541	0.000	0.000

Appendix 8 continued..

Reported in:	Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
	Cc02_g37350	Putative 1-aminocyclopropane-1-carboxylate oxidase homolog 1	0.125	GO:0016491	-2.436	0.000	0.000
	Cc02_g37750	myb domain protein 17	0.011	GO:0045449	-3.462	0.000	0.000
	Cc02_g38120	Aquaporin PIP1-3	0.046	GO:0005215	-4.581	0.000	0.000
Table 8	Cc02_g38150	Cytochrome P450 86A2	0.013	GO:0004497	-4.022	0.000	0.000
	Cc02_g38920	Putative glucose-6-phosphate 1-epimerase	0.017	GO:0005975	-3.024	0.000	0.000
	Cc02_g39830	Putative Transcription factor bHLH14	0.005	GO:0005634	-3.385	0.001	0.000
	Cc02_g39890	3-ketoacyl-CoA synthase 1	1.019	GO:0006633	-2.431	0.000	0.000
	Cc03_g00590	Invertase inhibitor	0.061	GO:0004857	-4.076	0.000	0.000
	Cc03_g01960	Putative Peroxidase 5	0.015	GO:0055114	-5.146	0.000	0.000
	Cc03_g02280	CASP-like protein VIT_05s0020g01820	0.775	N/A	-2.802	0.000	0.000
	Cc03_g02360	Probable serine/threonine-protein kinase At1g54610	0.006	GO:0004674	-2.094	0.003	0.000
	Cc03_g02460	Superoxide dismutase [Cu-Zn]	0.155	GO:0046872	-3.754	0.000	0.000
	Cc03_g02620	inosine-uridine preferring nucleoside hydrolase family protein	0.052	N/A	-3.718	0.000	0.000
	Cc03_g03580	Linoleate 9S-lipoxygenase 5, chloroplastic	0.022	GO:0005506	-9.670	0.000	0.000
	Cc03_g03600	Putative Uncharacterized protein C24B11.05	0.182	GO:0008152	-3.843	0.000	0.000
	Cc03_g04370	Protein argonaute 10	0.017	GO:0003676	-2.141	0.001	0.000
	Cc03_g04550	3-ketoacyl-CoA synthase 17	0.002	GO:0006633	-4.147	0.000	0.000
Table 8	Cc03_g04570	3-ketoacyl-CoA synthase 17	0.133	GO:0006633	-4.563	0.000	0.000
	Cc03_g05100	Protein of unknown function (DUF668)	0.038	N/A	-3.022	0.000	0.000
	Cc03_g05360	Putative Auxin-induced protein 5NG4	0.033	GO:0016020	-2.420	0.003	0.000
	Cc03_g05860	Hypothetical protein	0.038	N/A	-13.724	0.000	0.000
	Cc03_g05960	Peptide transporter PTR3-A	0.294	GO:0006857	-2.318	0.000	0.000
	Cc03_g07210	Receptor-like protein kinase 2	0.003	GO:0004674	-3.856	0.006	0.000
	Cc03_g08010	Hypothetical protein	0.011	N/A	134.866	0.005	0.000
	Cc03_g08020	Putative Cytosolic sulfotransferase 5	0.007	GO:0008146	89.625	0.001	0.000
	Cc03_g08690	Probable LRR receptor-like serine/threonine-protein kinase At4g08850	0.004	GO:0004674	-40.322	0.005	0.000
	Cc03_g11110	Hypothetical protein	2.106	N/A	-2.792	0.000	0.000
	Cc03_g12220	Desiccation-related protein PCC13-62	0.564	N/A	-2.407	0.000	0.000

Appendix 8 continued...

Reported in:	Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
	Cc03_g12230	Desiccation-related protein PCC13-62	2.866	N/A	-2.730	0.000	0.000
	Cc03_g12250	Desiccation-related protein PCC13-62	0.198	N/A	-2.773	0.000	0.000
	Cc03_g12260	Desiccation-related protein PCC13-62	0.223	N/A	-2.642	0.000	0.000
	Cc03_g13900	Putative Beta-amyrin 24-hydroxylase	0.010	GO:0004497	165.993	0.002	0.000
	Cc03_g14010	Putative Basic 7S globulin 2	0.014	GO:0006508	-3.097	0.000	0.000
	Cc04_g00680	Pectinesterase/pectinesterase inhibitor U1	0.360	GO:0005618	-2.616	0.000	0.000
	Cc04_g00860	Sucrose synthase	0.010	GO:0009058	-4.291	0.006	0.000
	Cc04_g01230	Cytochrome P450 76C2	0.011	GO:0004497	-2.279	0.005	0.000
	Cc04_g01540	Putative Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	0.039	N/A	-3.748	0.003	0.000
	Cc04_g01990	Beta-galactosidase	0.066	GO:0003824	-2.446	0.000	0.000
	Cc04_g02020	3-epi-6-deoxocathasterone 23-monooxygenase	0.031	GO:0004497	-2.048	0.000	0.000
	Cc04_g02930	Probable peptide transporter At1g52190	0.012	GO:0006857	-3.899	0.000	0.000
	Cc04_g03300	Putative Acid phosphatase 1	0.077	GO:0003993	2.477	0.002	0.000
	Cc04_g05080	Probable WRKY transcription factor 40	0.236	GO:0003700	-2.355	0.000	0.000
	Cc04_g05800	Sec14p-like phosphatidylinositol transfer family protein	0.019	N/A	-2.034	0.000	0.000
	Cc04_g08110	Tropinone reductase homolog At1g07440	0.023	GO:0055114	-2.964	0.003	0.000
	Cc04_g08330	Miraculin	0.013	GO:0004866	-5.551	0.000	0.000
	Cc04_g09620	Hypothetical protein	0.017	N/A	-7.790	0.000	0.000
	Cc04_g10480	Adenine/guanine permease AZG2	0.039	GO:0005215	-3.398	0.001	0.000
	Cc04_g10790	Vignain	0.087	GO:0006508	-2.765	0.002	0.000
	Cc04_g10940	Putative lipid transfer protein 1	0.006	GO:0008289	-45.818	0.000	0.000
	Cc04_g12810	Probable cyclic nucleotide-gated ion channel 16	0.024	GO:0005216	-2.037	0.000	0.000
	Cc04_g13440	Predicted protein	0.025	N/A	-5.888	0.000	0.000
	Cc04_g14360	Putative Receptor-like protein 12	0.005	GO:0005515	-2.327	0.001	0.000
	Cc04_g15770	Probable peptide/nitrate transporter At5g62680	0.276	GO:0006857	-3.189	0.000	0.000
	Cc04_g16290	NAC domain-containing protein 100	0.199	GO:0045449	-2.721	0.000	0.000

Appendix 8 continued

Reported in:	Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
	Cc04_g17150	Tubulin alpha chain	0.031	GO:0051258	-4.256	0.007	0.000
	Cc05_g00760	Acidic endochitinase	0.018	GO:0003824	13.980	0.008	0.000
	Cc05_g01510	Putative Disease resistance-responsive (dirigent-like protein) family protein	0.067	N/A	191.152	0.003	0.000
	Cc05_g01970	Major facilitator superfamily protein	0.037	N/A	-2.433	0.000	0.000
	Cc05_g02420	Fasciclin-like arabinogalactan protein 7	0.021	N/A	-3.347	0.000	0.000
Table 8	Cc05_g03130	Endoglucanase 11	0.005	GO:0003824	-5.942	0.000	0.000
	Cc05_g03500	Putative Protein ASPARTIC PROTEASE IN GUARD CELL 1	0.083	GO:0006508	-6.141	0.000	0.000
Table 8	Cc05_g03840	Potassium transporter 5	0.072	GO:0006813	-3.408	0.000	0.000
	Cc05_g04090	mitogen-activated protein kinase kinase kinase 18	0.044	GO:0004674	-2.869	0.000	0.000
	Cc05_g04110	Putative Disease resistance-responsive (dirigent-like protein) family protein	0.010	N/A	121.136	0.007	0.000
	Cc05_g04190	Putative Disease resistance-responsive (dirigent-like protein) family protein	0.029	N/A	59.654	0.004	0.000
	Cc05_g06040	Heat shock 70 kDa protein 4	0.007	GO:0005524	-2.991	0.000	0.000
	Cc05_g06330	Putative Reticuline oxidase-like protein	0.002	GO:0050660	-8.470	0.000	0.000
	Cc05_g06740	Probable protein phosphatase 2C 52	0.041	GO:0003824	-2.004	0.000	0.000
	Cc05_g06850	Protochlorophyllide reductase, chloroplastic	0.034	GO:0016630	-3.164	0.000	0.000
	Cc05_g08120	TRICHOME BIREFRINGENCE-LIKE 39	0.111	N/A	-2.869	0.000	0.000
	Cc05_g08560	Brassinosteroid-regulated protein BRU1	0.026	GO:0005618	-6.599	0.000	0.000
	Cc05_g08750	Secologanin synthase	0.018	GO:0004497	-4.390	0.000	0.000
	Cc05_g08980	Putative G-type lectin S-receptor-like serine/threonine-protein kinase SD2-5	0.335	GO:0004674	-2.405	0.000	0.000
	Cc05_g08990	Putative G-type lectin S-receptor-like serine/threonine-protein kinase At5g35370	8.916	GO:0005529	-2.369	0.000	0.000
	Cc05_g09270	Probable rhamnose biosynthetic enzyme 1	0.007	GO:0050662	-4.629	0.000	0.000
	Cc05_g09500	Glycerol-3-phosphate acyltransferase 1	0.245	GO:0008415	-2.941	0.000	0.000
	Cc05_g09630	Putative Probable WRKY transcription factor 71	0.053	GO:0003700	-2.720	0.000	0.000
	Cc05_g10050	Putative Protein kinase G11A	0.004	GO:0004674	-2.891	0.007	0.000

Appendix 8 continued…

Reported in:	Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
	Cc05_g10310	Polyphenol oxidase I, chloroplastic	0.064	GO:0055114	-6.833	0.000	0.000
	Cc05_g11140	Putative Epidermis-specific secreted glycoprotein EP1	0.092	GO:0005529	-12.040	0.000	0.000
	Cc05_g11240	TPR repeat-containing thioRedoxin TTL1	0.023	GO:0045454	-2.122	0.004	0.000
	Cc05_g12490	ABC transporter G family member 9	0.022	GO:0016020	-4.096	0.000	0.000
	Cc05_g15530	Probable galactinol--sucrose galactosyltransferase 1	0.012	N/A	-2.378	0.000	0.000
	Cc06_g01510	Auxin transporter-like protein 2	0.094	N/A	-3.352	0.000	0.000
	Cc06_g02310	Putative Protein of unknown function (DUF668)	0.015	N/A	-2.154	0.005	0.000
	Cc06_g02510	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein	1.397	GO:0016787	-2.110	0.000	0.000
	Cc06_g02880	Short-chain dehydrogenase TIC 32, chloroplastic	0.036	GO:0055114	-2.567	0.003	0.000
	Cc06_g03720	L-type lectin-domain containing receptor kinase VIII.1	0.028	GO:0004674	-2.007	0.000	0.000
	Cc06_g03900	Aquaporin PIP2-2	0.035	GO:0005215	-2.949	0.000	0.000
	Cc06_g04240	Mannan synthase 1	1.986	N/A	-2.569	0.000	0.000
	Cc06_g04490	Putative Bifunctional monodehydroascorbate reductase and carbonic anhydrase nectarin-3	0.009	GO:0006730	-3.378	0.000	0.000
Table 8	Cc06_g04690	Putative pectinesterase 14	0.004	GO:0005618	-19.223	0.000	0.000
	Cc06_g04760	Putative RING-H2 finger protein ATL52	0.004	GO:0008270	-4.083	0.007	0.000
	Cc06_g04800	Pollen Ole e 1 allergen and extensin family protein	0.011	N/A	-2.820	0.001	0.000
	Cc06_g04920	Beta-galactosidase	0.001	GO:0003824	-9.941	0.000	0.000
	Cc06_g07510	Putative uncharacterized protein	0.070	N/A	3.225	0.000	0.000
	Cc06_g08460	Peroxidase 25	0.046	GO:0055114	-4.674	0.000	0.000
	Cc06_g09370	Putative Sigma factor sigB regulation protein rsbQ	0.060	N/A	-3.311	0.000	0.000
	Cc06_g09500	Hypothetical protein	0.008	N/A	-3.041	0.007	0.000
	Cc06_g09700	Polyol transporter 5	0.004	GO:0022891	-14.833	0.000	0.000
	Cc06_g09810	Putative unknown protein; FUNCTIONS IN	0.030	N/A	-2.909	0.000	0.000

Appendix 8 continued…

Reported in:	Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
	Cc06_g09840	Putative Dynein light chain LC6, flagellar outer arm	0.022	GO:0007017	2.768	0.000	0.000
	Cc06_g11710	unknown protein; BEST Arabidopsis thaliana protein match is	0.041	N/A	-2.001	0.000	0.000
	Cc06_g11870	Protein Brevis radix-like 4	0.023	N/A	-2.057	0.000	0.000
	Cc06_g12250	Putative Protein of unknown function DUF567	0.127	N/A	-2.317	0.009	0.000
	Cc06_g13390	CASP-like protein 3	0.059	N/A	-2.482	0.000	0.000
	Cc06_g13430	Probable xyloglucan endotransglucosylase/hydrolase protein 23	0.043	GO:0005618	-2.785	0.000	0.000
Table 8	Cc06_g13440	Xyloglucan endotransglucosylase/hydrolase protein 22	0.457	GO:0005618	-4.688	0.000	0.000
	Cc06_g13710	Predicted protein	0.421	N/A	-2.353	0.000	0.000
	Cc06_g13790	BNR/Asp-box repeat family protein	0.008	N/A	-2.427	0.000	0.000
	Cc06_g14080	Protein WAX2	0.207	GO:0006633	-2.573	0.003	0.000
	Cc06_g14410	Invertase inhibitor	0.006	N/A	-10.112	0.006	0.000
	Cc06_g15740	Homeobox-leucine zipper protein HAT14	0.012	GO:0043565	-2.345	0.002	0.000
	Cc06_g15760	Peroxidase 15	0.004	GO:0055114	-9.699	0.000	0.000
	Cc06_g16270	Polygalacturonase inhibitor	0.453	GO:0005515	-2.469	0.000	0.000
	Cc06_g16280	Polygalacturonase inhibitor	8.911	GO:0005515	-2.605	0.000	0.000
	Cc06_g16290	Polygalacturonase inhibitor	0.356	GO:0005515	-3.068	0.000	0.000
	Cc06_g16560	Potassium transporter 5	0.016	GO:0006813	-2.435	0.000	0.000
	Cc06_g17700	Agamous-like MADS-box protein AGL15	0.033	GO:0003700	-4.064	0.000	0.000
	Cc06_g18090	6-phosphogluconate dehydrogenase family protein	0.053	GO:0004367	-2.547	0.000	0.000
	Cc06_g18380	unknown protein; BEST Arabidopsis thaliana protein match is	0.069	N/A	-3.680	0.002	0.000
	Cc06_g18440	Non-specific lipid-transfer protein	15.310	GO:0008289	-3.032	0.000	0.000
	Cc06_g20950	Putative (S)-N-methylcoclaurine 3'-hydroxylase isozyme 2	0.002	GO:0004497	-6.678	0.000	0.000
	Cc06_g21200	Metacaspase-9	0.082	GO:0006508	-17.613	0.000	0.000
	Cc06_g21680	Cadmium/zinc-transporting ATPase HMA2	0.032	GO:0016787	-2.529	0.002	0.000

Appendix 8 continued…

Reported in:	Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
	Cc06_g22410	Hypothetical protein	0.084	N/A	70.967	0.008	0.000
	Cc06_g22750	Probable S-acyltransferase At5g05070	0.011	GO:0008270	-2.857	0.000	0.000
	Cc06_g22850	Short-chain type dehydrogenase/reductase	0.861	GO:0055114	-2.014	0.000	0.000
	Cc06_g22980	FUNCTIONS IN	0.043	N/A	-2.243	0.000	0.000
	Cc06_g23480	Hypothetical protein	0.050	GO:0005515	-6.002	0.000	0.000
	Cc07_g00100	Cytochrome P450 724B1	0.039	GO:0004497	-3.967	0.000	0.000
	Cc07_g00690	Putative Mavicyanin	0.023	GO:0005507	-2.530	0.005	0.000
	Cc07_g02770	Probable galacturonosyltransferase-like 9	0.014	N/A	-2.329	0.000	0.000
	Cc07_g03350	Myb-related protein 306	0.769	GO:0045449	-2.409	0.000	0.000
	Cc07_g03490	Expansin-A15	0.232	GO:0009664	-3.448	0.001	0.000
	Cc07_g03560	Casparian strip membrane protein POPTRDRAFT_569472	0.038	N/A	-3.018	0.000	0.000
	Cc07_g03940	4-coumarate--CoA ligase-like 1	0.172	GO:0003824	-5.062	0.000	0.000
	Cc07_g04430	Putative myb domain protein 121	0.010	GO:0045449	-4.628	0.002	0.000
	Cc07_g05130	Putative Protein of unknown function (DUF607)	0.086	N/A	-2.411	0.000	0.000
	Cc07_g05140	Hypothetical protein	0.177	N/A	-2.271	0.000	0.000
	Cc07_g05590	Putative Probable esterase PIR7A	0.138	N/A	-2.843	0.000	0.000
	Cc07_g05670	L-ascorbate oxidase homolog	0.132	GO:0005507	-3.149	0.000	0.000
	Cc07_g06680	Arabinogalactan	0.664	N/A	-3.489	0.000	0.000
	Cc07_g06750	Subtilisin-like protease	0.258	GO:0004252	-2.408	0.002	0.000
	Cc07_g07030	Calcium-transporting ATPase 2, plasma membrane-type	0.042	GO:0006812	-2.181	0.000	0.000
	Cc07_g07540	Homeobox-leucine zipper protein HAT22	0.050	GO:0043565	-2.276	0.000	0.000
	Cc07_g07790	Uncharacterized protein At5g65660	0.169	N/A	-2.184	0.000	0.000
	Cc07_g08030	Transcription factor MYB113	3.706	GO:0045449	2.040	0.001	0.000
	Cc07_g08090	Protein kinase superfamily protein	0.044	GO:0004674	-3.117	0.000	0.000
	Cc07_g08370	Putative unknown protein; FUNCTIONS IN	0.020	N/A	-2.231	0.004	0.000
	Cc07_g08490	Putative uncharacterized protein	0.229	N/A	-2.270	0.000	0.000
	Cc07_g08550	Probable non-specific lipid-transfer protein AKCS9	2.218	GO:0006869	-2.903	0.000	0.000
	Cc07_g09440	Cytochrome P450 90B1	0.073	GO:0004497	-2.130	0.003	0.000

Appendix 8 continued…

Reported in:	Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
	Cc07_g10200	Beta-galactosidase 3	0.091	GO:0003824	-4.336	0.000	0.000
	Cc07_g10730	Putative Snakin-2	0.049	N/A	-3.486	0.000	0.000
	Cc07_g11860	Putative Transducin/WD40 repeat-like superfamily protein	0.046	N/A	-2.011	0.001	0.000
	Cc07_g12360	Putative Probable L-type lectin-domain containing receptor kinase S.5	0.022	GO:0004674	-2.285	0.006	0.000
	Cc07_g13250	Eukaryotic aspartyl protease family protein	0.021	GO:0006508	-2.619	0.000	0.000
	Cc07_g13390	Glucan endo-1,3-beta-glucosidase 2	0.002	GO:0003824	-3.873	0.003	0.000
	Cc07_g13620	Putative Rhamnogalacturonate lyase	0.644	GO:0003824	3.595	0.000	0.000
	Cc07_g13930	Putative UDP-rhamnose	0.245	GO:0008152	-3.452	0.000	0.000
	Cc07_g14130	3-hydroxyacyl-CoA dehydratase PASTICCINO 2	0.202	N/A	-2.275	0.000	0.000
Table 8	Cc07_g14260	Putative Probable mitochondrial 2-oxoglutarate/malate carrier protein	0.397	GO:0016020	-2.242	0.000	0.000
	Cc07_g14270	unknown protein; FUNCTIONS IN	0.060	N/A	-2.189	0.000	0.000
	Cc07_g14580	unknown seed protein like 1	0.234	N/A	-2.061	0.000	0.000
	Cc07_g16130	Putative Cytochrome P450 81D1	0.014	GO:0004497	164.216	0.006	0.000
	Cc07_g20210	Probable aldo-keto reductase 2	0.058	GO:0055114	-2.988	0.000	0.000
	Cc07_g20370	Alcohol dehydrogenase-like 2	0.006	GO:0008270	-4.830	0.000	0.000
	Cc07_g21010	Putative peptide/nitrate transporter At2g38100	0.008	GO:0006857	-6.272	0.000	0.000
	Cc07_g21360	Hypothetical protein	0.034	GO:0003700	-2.052	0.000	0.000
	Cc08_g00740	Ethylene-responsive transcription factor SHINE 3	0.006	GO:0003700	-8.036	0.002	0.000
	Cc08_g01290	Thermospermine synthase ACAULIS5	0.005	GO:0003824	-3.851	0.001	0.000
	Cc08_g02100	Two-component response regulator-like APRR2	0.033	GO:0000160	-2.039	0.000	0.000
	Cc08_g02850	Hypothetical protein	0.011	N/A	-7.108	0.000	0.000
	Cc08_g02900	Tetraspanin family protein	0.003	GO:0016021	-4.373	0.002	0.000
	Cc08_g03030	ABC transporter B family member 2	2.218	GO:0016021	-2.244	0.000	0.000
	Cc08_g04330	Isoflavone reductase homolog	0.008	GO:0003824	-4.380	0.000	0.000
	Cc08_g04340	Isoflavone reductase homolog	0.207	GO:0003824	-4.196	0.000	0.000

Appendix 8 continued…

Reported in:	Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
	Cc08_g04630	Pectinacetylesterase family protein	0.131	N/A	-2.301	0.000	0.000
	Cc08_g06740	Putative Aspartic proteinase nepenthesin-2	0.498	GO:0006508	-7.990	0.000	0.000
	Cc08_g06810	Protein of unknown function, DUF599	0.018	N/A	-21.802	0.000	0.000
	Cc08_g07460	Premnaspirodiene oxygenase	0.094	GO:0004497	-2.260	0.001	0.000
	Cc08_g10360	Putative Probable protein phosphatase 2C 74	0.077	GO:0003824	-2.106	0.001	0.000
	Cc08_g11080	Hypothetical protein	0.692	N/A	-2.087	0.000	0.000
	Cc08_g11520	Putative HXXXD-type acyl-transferase family protein	0.021	N/A	-3.908	0.000	0.000
	Cc08_g12010	Putative Subtilisin-like protease SDD1	0.002	GO:0004252	-23.425	0.003	0.000
	Cc08_g12720	L-lactate dehydrogenase A	0.016	GO:0005737	-4.299	0.000	0.000
	Cc08_g12800	Nitrate transporter 1.1	0.113	GO:0006857	-3.460	0.000	0.000
	Cc08_g12840	Adenine phosphoribosyltransferase 1, chloroplastic	0.028	GO:0005737	-5.025	0.000	0.000
	Cc08_g12940	Monocopper oxidase-like protein SKU5	0.016	GO:0005507	-6.051	0.000	0.000
	Cc08_g13590	Putative Cytochrome P450 76C4	0.007	GO:0004497	-2.447	0.000	0.000
	Cc08_g13880	Protein HOTHEAD	0.029	GO:0006066	-3.071	0.000	0.000
	Cc08_g13950	Putative unknown protein; FUNCTIONS IN	0.011	N/A	-2.434	0.000	0.000
	Cc08_g14470	Putative O-glucosyltransferase rumi homolog	0.016	N/A	-2.588	0.000	0.000
	Cc08_g14520	Putative copper-transporting ATPase HMA5	0.019	GO:0004008	-2.399	0.000	0.000
	Cc08_g14880	Cytochrome P450 86A2	0.029	GO:0004497	-3.667	0.000	0.000
	Cc08_g15660	Putative glucose-6-phosphate 1-epimerase	0.203	GO:0005975	2.195	0.000	0.000
	Cc08_g16800	alpha/beta-Hydrolases superfamily protein	0.038	N/A	-4.303	0.000	0.000
	Cc08_g16810	GDSL esterase/lipase At5g41890	0.006	GO:0006629	-8.494	0.000	0.000
	Cc08_g16920	Alcohol dehydrogenase-like 3	0.011	GO:0008270	-2.744	0.000	0.000
	Cc09_g00220	Farnesyl pyrophosphate synthase 1	0.021	GO:0008299	-3.488	0.000	0.000
	Cc09_g00330	Leafy cotyledon 1-like protein	0.008	GO:0043565	-3.860	0.004	0.000
	Cc09_g02970	Leucine-rich repeat receptor-like protein kinase PXL1	0.001	GO:0004674	-2.673	0.007	0.000
	Cc09_g03070	Protein of Unknown Function (DUF239)	0.012	N/A	-3.617	0.000	0.000
	Cc09_g03110	GDSL esterase/lipase At5g33370	0.009	GO:0006629	-7.155	0.007	0.000
	Cc09_g03380	Protein of unknown function, DUF538	12.507	N/A	-2.199	0.000	0.000

Appendix 8 continued…

Reported in:	Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
	Cc09_g03770	Serine carboxypeptidase-like 20	0.006	GO:0006508	-21.756	0.005	0.000
	Cc09_g04350	L-ascorbate oxidase homolog	0.015	GO:0005507	-2.720	0.000	0.000
	Cc09_g04850	Hypothetical protein	0.157	N/A	154.490	0.002	0.000
	Cc09_g05150	Putative Cytochrome P450 89A2	0.005	GO:0004497	-5.969	0.000	0.000
	Cc09_g05430	Hypothetical protein	1.289	N/A	-50.110	0.000	0.000
	Cc09_g06260	Putative Protein PHLOEM PROTEIN 2-LIKE A9	0.256	N/A	-2.627	0.000	0.000
	Cc09_g06270	Putative Protein PHLOEM PROTEIN 2-LIKE A9	0.016	N/A	-3.903	0.000	0.000
	Cc09_g06970	7-methylxanthosine synthase 1	1.875	GO:0008168	-3.316	0.000	0.000
	Cc09_g07170	Putative Reticuline oxidase-like protein	0.001	GO:0050660	-7.318	0.000	0.000
	Cc09_g07180	Putative Reticuline oxidase-like protein	0.006	GO:0050660	-7.978	0.000	0.000
	Cc09_g07720	Hypothetical protein	0.550	N/A	2.329	0.003	0.000
	Cc09_g08540	Phosphate-responsive 1 family protein	0.029	N/A	-2.704	0.000	0.000
	Cc09_g09160	Probable purine permease 9	0.005	N/A	-3.657	0.002	0.000
	Cc09_g09270	Putative Ankyrin repeat family protein	0.002	GO:0006952	-2.911	0.006	0.000
	Cc09_g09780	Putative uncharacterized protein	0.005	GO:0008270	-8.003	0.000	0.000
	Cc09_g10100	Putative Subtilisin-like protease	0.052	GO:0004252	235.900	0.000	0.000
	Cc09_g10270	Putative uncharacterized protein	0.011	N/A	-2.528	0.001	0.000
	Cc09_g10800	EXORDIUM like 5	0.239	N/A	-3.148	0.000	0.000
	Cc10_g00910	L-ascorbate oxidase homolog	0.499	GO:0005507	-3.167	0.000	0.000
	Cc10_g01310	Probable cinnamyl alcohol dehydrogenase 6	0.010	GO:0008270	-5.568	0.000	0.000
	Cc10_g01820	Endoglucanase 24	0.015	GO:0003824	-5.355	0.000	0.000
	Cc10_g02330	Beta-galactosidase 3	0.008	GO:0003824	-2.358	0.000	0.000
	Cc10_g02450	Putative Transmembrane protein 136	0.015	GO:0016021	-3.443	0.000	0.000
	Cc10_g03370	Pathogenesis-related protein 5	0.007	N/A	-4.683	0.000	0.000
	Cc10_g03880	Protein of unknown function, DUF617	0.076	N/A	-2.676	0.004	0.000
	Cc10_g03960	Serine carboxypeptidase-like 40	0.003	GO:0006508	-9.467	0.000	0.000
	Cc10_g04170	Alpha-amylase	0.661	GO:0003824	-2.319	0.000	0.000
	Cc10_g04580	Putative Subtilisin-like protease	0.027	GO:0004252	-2.557	0.003	0.000
	Cc10_g04910	Probable carbohydrate esterase At4g34215	0.012	N/A	-2.322	0.009	0.000

Appendix 8 continued…

Reported in:	Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
	Cc10_g05460	Snakin-1	0.003	N/A	-10.488	0.002	0.000
	Cc10_g06690	Putative Homeobox-leucine zipper protein ATHB-16	0.011	GO:0043565	-3.080	0.000	0.000
	Cc10_g06820	Phospho-2-dehydro-3-deoxyheptonate aldolase 1, chloroplastic	0.648	GO:0003849	-2.004	0.000	0.000
	Cc10_g06840	Phosphate-responsive 1 family protein	0.012	N/A	-2.981	0.002	0.000
	Cc10_g06870	Phosphate-responsive 1 family protein	0.294	N/A	-2.494	0.008	0.000
	Cc10_g07410	Probable trehalose-phosphate phosphatase E	0.008	GO:0003824	-3.413	0.000	0.000
	Cc10_g07480	Probable beta-D-xylosidase 7	0.012	GO:0005975	-5.291	0.000	0.000
	Cc10_g07630	Putative uncharacterized protein	0.061	N/A	2.825	0.000	0.000
	Cc10_g08680	Probable peptide/nitrate transporter At3g54450	0.309	GO:0006857	-3.315	0.000	0.000
	Cc10_g08840	Metallothionein-like protein 4A	0.202	GO:0046872	-4.022	0.003	0.000
	Cc10_g09390	Agamous-like MADS-box protein AGL11	0.364	GO:0003700	-2.382	0.000	0.000
	Cc10_g09590	Hypothetical protein	0.091	N/A	-5.141	0.000	0.000
	Cc10_g10920	Putative Auxin-induced protein 5NG4	1.652	GO:0016020	-2.549	0.000	0.000
	Cc10_g10960	Putative Ethylene-responsive transcription factor ERF014	0.020	GO:0003700	-3.148	0.000	0.000
	Cc10_g11070	Flavonol synthase/flavanone 3-hydroxylase	0.056	GO:0016491	-5.031	0.000	0.000
	Cc10_g11300	Serine/threonine-protein kinase-like protein CCR4	0.001	GO:0004674	-7.500	0.000	0.000
	Cc10_g12710	1-aminocyclopropane-1-carboxylate oxidase 5	0.022	GO:0016491	-18.791	0.001	0.000
	Cc10_g12840	Hypothetical protein	0.017	N/A	5.054	0.001	0.000
	Cc10_g15770	Putative Protein ASPARTIC PROTEASE IN GUARD CELL 1	0.017	GO:0006508	-5.194	0.000	0.000
	Cc10_g16350	Putative DNA-damage-repair/toleration protein DRT100	0.146	GO:0005515	-2.240	0.000	0.000
Table 8	Cc10_g16390	GDSL esterase/lipase At1g71250	0.322	N/A	2.251	0.002	0.000
	Cc10_g16520	Putative 12-oxophytodienoate reductase 11	0.008	GO:0016491	-3.501	0.000	0.000
	Cc11_g01870	Putative Acyl-protein thioesterase 1 homolog 1	0.044	GO:0016787	-3.126	0.000	0.000

Appendix 8 continued…

Reported in:	Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
	Cc11_g02660	Putative Probable LRR receptor-like serine/threonine-protein kinase At3g47570	0.024	GO:0004674	-2.078	0.000	0.000
	Cc11_g05310	G-type lectin S-receptor-like serine/threonine-protein kinase SD2-5	0.002	GO:0004674	-4.890	0.002	0.000
	Cc11_g05350	TRAM, LAG1 and CLN8 (TLC) lipid-sensing domain containing protein	0.011	N/A	-4.635	0.000	0.000
	Cc11_g05750	Putative ABC transporter B family member 12	0.000	GO:0016021	-8.995	0.002	0.000
	Cc11_g05930	Putative Probable transmembrane ascorbate ferrireductase 3	0.293	GO:0016021	-2.731	0.007	0.000
	Cc11_g06610	D-arabinono-1,4-lactone oxidase family protein	0.010	GO:0016020	-11.005	0.000	0.000
	Cc11_g07040	Nudix hydrolase 17, mitochondrial	0.005	GO:0016787	-4.441	0.000	0.000
	Cc11_g07130	Lignin-forming anionic peroxidase	0.005	GO:0055114	-5.457	0.001	0.000
	Cc11_g07770	Pathogenesis-related protein R major form	0.031	N/A	-6.853	0.000	0.000
	Cc11_g08520	Probable sulfate transporter 3.3	0.009	GO:0005215	-3.975	0.000	0.000
	Cc11_g08680	Auxin efflux carrier component 3	0.085	GO:0016021	-2.062	0.000	0.000
	Cc11_g08850	Putative Spore coat protein A	0.041	GO:0005507	-2.206	0.007	0.000
	Cc11_g09370	Acidic endochitinase	0.003	GO:0003824	-18.912	0.000	0.000
	Cc11_g10000	Endonuclease 2	0.158	GO:0004519	-2.119	0.004	0.000
	Cc11_g10030	Adenylate isopentenyltransferase	0.081	GO:0004811	-2.912	0.000	0.000
	Cc11_g10180	3-ketoacyl-CoA synthase 6	0.004	GO:0006633	-3.355	0.000	0.000
	Cc11_g10230	Alpha-xylosidase	0.021	GO:0005975	-2.046	0.007	0.000
	Cc11_g10400	Probable carboxylesterase 17	0.122	GO:0016787	-3.448	0.000	0.000
	Cc11_g10530	C2H2-like zinc finger protein	0.021	GO:0008270	-4.650	0.005	0.000
	Cc11_g10560	AP2/ERF and B3 domain-containing transcription repressor TEM1	0.109	GO:0003700	-4.003	0.000	0.000
	Cc11_g10580	Probable galacturonosyltransferase-like 3	0.017	N/A	-3.492	0.000	0.000
	Cc11_g11970	Ureide permease 2	0.110	N/A	-2.116	0.000	0.000
	Cc11_g12120	14-3-3-like protein GF14 iota	0.012	GO:0019904	-9.287	0.000	0.000
	Cc11_g12890	Putative Homeodomain-like superfamily protein	0.004	GO:0045449	-4.990	0.000	0.000

Appendix 8 continued…

Reported in:	Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
	Cc11_g12930	Extracellular ribonuclease LE	0.015	GO:0033897	-2.581	0.001	0.000
	Cc11_g13720	HIPL1 protein	0.037	GO:0003824	-2.597	0.001	0.000
	Cc11_g14010	Probable galacturonosyltransferase-like 9	0.064	N/A	-2.208	0.000	0.000
	Cc11_g14320	Nucleobase-ascorbate transporter 7	0.012	GO:0005215	-2.484	0.000	0.000
	Cc11_g14370	MLP-like protein 423	3.944	GO:0006952	-8.397	0.000	0.000
	Cc11_g14690	Polygalacturonase-1 non-catalytic subunit beta	0.053	N/A	-2.980	0.000	0.000
	Cc11_g15550	Putative protein kinase family protein / peptidoglycan-binding LysM domain-containing protein	0.008	GO:0004674	-3.122	0.000	0.000
	Cc11_g16970	Putative Uncharacterized GPI-anchored protein At1g27950	0.082	N/A	-3.792	0.000	0.000
	Cc11_g17250	Putative Predicted protein	0.004	N/A	-4.632	0.009	0.000
	Cc11_g17390	Putative Bidirectional sugar transporter SWEET15	1.336	GO:0016021	-13.159	0.000	0.000

APPENDIX 9

Unique and shared immature DEGs between the cultivars and within treatments. This shows some shared DEGs across treatments from the Venn diagrams in Figure 20.

Treatments	Inbred		Hybrid	
	Shared	DEG	Shared	DEG
C; R; RT; T; Total immature	1	Respiratory burst oxidase homolog protein B	3	Cytochrome P450 86A2; Potassium transporter 5; 3-ketoacyl-CoA synthase 17
T; Total immature	9	NAC domain-containing protein 100; Nitrate transporter 1.1; ABC transporter G family member 9; D-arabinono-1,4-lactone oxidase family protein; Beta-xylosidase/alpha-L-arabinofuranosidase 2; Momilactone A synthase; Tropinone reductase homolog At1g07440; Putative uncharacterized protein; Protein WAX2	24	Alcohol dehydrogenase-like 2; Probable purine permease 9; Alpha-xylosidase; Heat shock 70 kDa protein 4; Aquaporin PIP2-2; Putative Probable LRR receptor-like serine/threonine-protein kinase At3g47570; Putative 12-oxophytodienoate reductase 11; Polygalacturonase inhibitor; Lignin-forming anionic peroxidase; Phosphate-responsive 1 family protein; unknown seed protein like 1; Nucleobase-ascorbate transporter 7; Probable rhamnose biosynthetic enzyme 1; Alcohol dehydrogenase-like 3; Protein argonaute 10; Beta-galactosidase 3; Probable histone H2B.1; Thermospermine synthase ACAULIS5; Serine carboxypeptidase-like 48; Glucan endo-1,3-beta-glucosidase 7; Acidic endochitinase; Probable polygalacturonase; Sulfate transporter 3.1; Aquaporin TIP1-1
C; Total immature	10	Uncharacterized protein At5g65660; Endoglucanase 12; Farnesyl pyrophosphate synthase 1; Probable pectinesterase/pectinesterase inhibitor 12; TRICHOME BIREFRINGENCE-LIKE 39; L-lactate dehydrogenase A; Putative Protein of unknown function (DUF668); Beta-galactosidase; Pectinacetylesterase family protein; Polyphenol oxidase I, chloroplastic	9	Homeobox-leucine zipper protein HAT14; Mannan synthase 1 6-phosphogluconate dehydrogenase family protein; Receptor-like protein kinase 2; Putative 3-ketoacyl-CoA synthase 5; Leafy cotyledon 1-like protein; Probable carbohydrate esterase At4g34215; MLP-like protein 423; Putative UDP-glycosyltransferase 85A2

Appendix 9 continued…

Treatments	Inbred		Hybrid	
	Shared	DEG	Shared	DEG
R; Total immature	1	Metallothionein-like protein 4A	37	Xyloglucan endotransglucosylase/hydrolase protein 22; Protein of unknown function (DUF668) mitogen-activated protein kinase kinase kinase 18; Putative HVA22-like protein g; Putative Probable NADH dehydrogenase; Probable cinnamyl alcohol dehydrogenase 6; Probable sulfate transporter 3.3; Pleiotropic drug resistance protein 3; Putative Probable transmembrane ascorbate ferrireductase 3; UDP-glycosyltransferase 85A2; Tetraspanin family protein; Protein of unknown function, DUF617; HIPL1 protein; Ureide permease 2; Probable non-specific lipid-transfer protein AKCS9; Probable peptide transporter At1g52190; L-type lectin-domain containing receptor kinase VIII.1; Protein kinase superfamily protein 3-ketoacyl-CoA synthase 6; Peptide transporter PTR3-A; SPX domain-containing protein 3; Putative Disease resistance response protein 206; Adenine/guanine permease AZG2; Putative Probable glucan 1,3-beta-glucosidase A; Putative Probable flavonol synthase 5 CASP-like protein 3; Adenine phosphoribosyltransferase 1, chloroplastic; Auxin transporter-like protein 2; GDSL esterase/lipase At5g41890; Linoleate 13S-lipoxygenase 2-1, chloroplastic; Putative Probable mitochondrial 2-oxoglutarate/malate carrier protein; Putative copper-transporting ATPase HMA5; unknown protein; BEST Arabidopsis thaliana protein match is myb domain protein 17 4-coumarate--CoA ligase-like 1; Putative Major facilitator superfamily domain-containing protein 12; Protein of unknown function, DUF599

APPENDIX 10

Significant GO terms obtained from the SEA analysis according to the maturity stage. Ontologies involved in molecular functions (F), cellular components (C), and biological process (P) categories, are described for each GO term. Input list corresponds to observed DEGs in this study, while numbers in reference belong to pre-reported for *Coffea canephora* [134].

Maturity	#	GO term	Ontology	Description	Number in input list	Number in reference	*p*-value	FDR
Immature	1	0055114	P	oxidation reduction	53	1242	6.0E-07	0.0002
Immature	2	0005975	P	carbohydrate metabolic process	38	795	2.4E-06	0.0005
Immature	3	0006633	P	fatty acid biosynthetic process	8	68	1.5E-04	0.0190
Immature	4	0006631	P	fatty acid metabolic process	8	77	3.2E-04	0.0310
Immature	5	0016491	F	oxidoreductase activity	68	1634	2.6E-08	0.0000
Immature	6	0003824	F	catalytic activity	191	7227	2.0E-06	0.0004
Immature	7	0005506	F	iron ion binding	31	685	5.9E-05	0.0068
Immature	8	0070001	F	aspartic-type peptidase activity	8	76	3.0E-04	0.0210
Immature	9	0004190	F	aspartic-type endopeptidase activity	8	76	3.0E-04	0.0210
Immature	10	0016701	F	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	5	26	3.7E-04	0.0210
Immature	11	0016762	F	xyloglucan:xyloglucosyl transferase activity	5	30	6.5E-04	0.0330

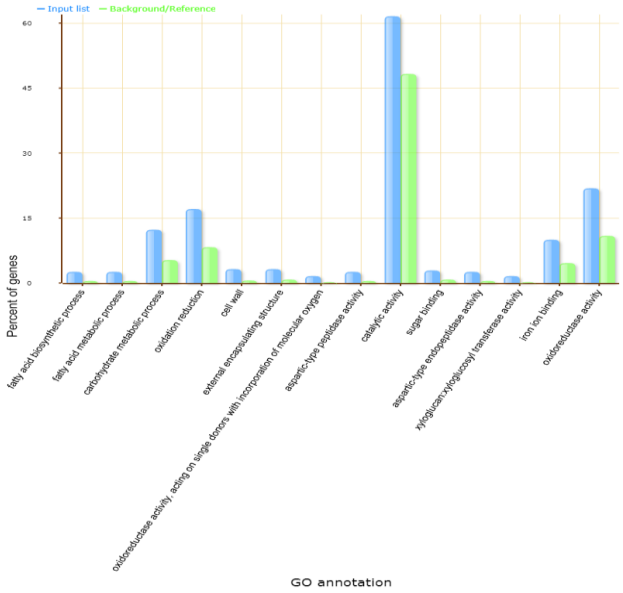
Appendix 10 continued…

Maturity	#	GO term	Ontology	Description	Number in input list	Number in reference	*p*-value	FDR
Immature	12	0005529	F	sugar binding	9	117	1.1E-03	0.0470
Immature	13	0005618	C	cell wall	10	96	5.8E-05	0.0039
Immature	14	0030312	C	external encapsulating structure	10	121	3.4E-04	0.0110
Mature	1	0008171	F	O-methyltransferase activity	6	61	1.7E-09	0.0000
Mature	2	0046983	F	protein dimerization activity	5	137	4.8E-06	0.0001
Mature	3	0016741	F	transferase activity, transferring one-carbon groups	6	282	1.0E-05	0.0001
Mature	4	0008168	F	methyltransferase activity	6	280	9.6E-06	0.0001
Mature	5	0003824	F	catalytic activity	20	7227	5.9E-03	0.0440

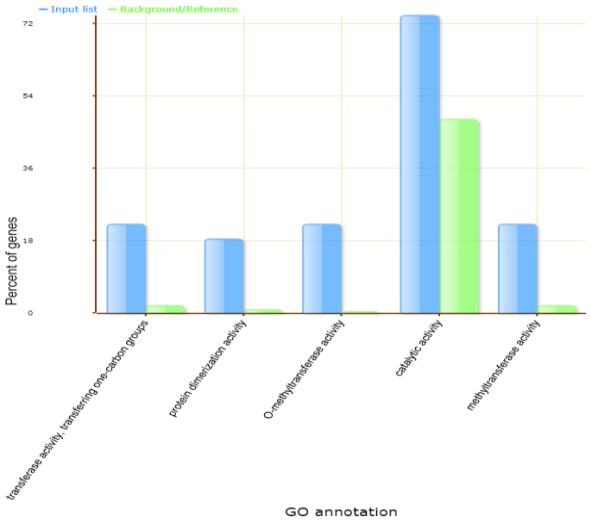
APPENDIX 11

Functional GO terms enriched by the DEGs between maturity stages in the hybrid and inbred cultivars.

A. Significant GO terms – Immature beans



B. Significant GO terms – Mature beans



A) Represents the 310 DEGs found in the SEA out of the 426 annotated genes, categorized in 14 GO terms. B) Represents the 27 DEGs found in the SEA out of the 45 annotated genes, categorized in 5 GO terms. The blue color bar in the chart represents the estimated value, while green represents the background/reference used to compare.

APPENDIX 12

Spearman correlation and significance between the green-bean volatiles and DEGs.

Cultivar	Maturity stage	Candidate gene (Gene ID [134])	Volatile	Spearman p	*p*-value	Functional GO term(s) associated
Inbred	Mature	GDSL esterase/lipase At1g71250 (Cc10_g16390)	Nonanal	0.72	0.0117	
Hybrid	Mature	GDSL esterase/lipase At1g71250 (Cc10_g16390)	Nonanal	0.81	0.0013	
Hybrid	Mature	GDSL esterase/lipase At1g71250 (Cc10_g16390)	(E)-2-Decenal	0.64	0.0247	
Hybrid	Immature	Probable cyclic nucleotide-gated ion channel 16 (Cc04_g12810)	dl-Limonene	0.75	0.0047	GO:0005216, GO:0016020, GO:0006811, GO:0055085
Inbred	Immature	Fasciclin-like arabinogalactan protein 8 (Cc02_g36510)	(E)-2-Decenal	0.61	0.0478	
Hybrid	Immature	Fasciclin-like arabinogalactan protein 8 (Cc02_g36510)	dl-Limonene	0.66	0.0193	
Hybrid	Mature	Putative Isoflavone 2'-hydroxylase (Cc00_g25270)	(E)-2-Decenal	0.74	0.0061	GO:0004497, GO:0020037, GO:0055114, GO:0009055
Inbred	Mature		(E)-2-Decenal	0.74	0.0090	
Hybrid	Mature	Putative Disease resistance-responsive (dirigent-like protein) family protein (Cc00_g04320)	(E)-2-Decenal	0.74	0.0061	
Hybrid	Immature	Endoglucanase 24 (Cc10_g01820)	dl-Limonene	0.71	0.0102	GO:0003824, GO:0005975

Appendix 12 continued…

Cultivar	Maturity stage	Candidate gene (Gene ID [134])	Volatile	Spearman p	*p*-value	Functional GO term(s) associated
Hybrid	Immature	Putative Domain of unknown function (DUF966) (Cc02_g26760)	Nonanal	-0.75	0.0053	
Hybrid	Immature	Putative Domain of unknown function (DUF966) (Cc02_g26760)	(E)-2-Decenal	-0.65	0.0220	
Hybrid	Immature	Putative Domain of unknown function (DUF966) (Cc02_g26760)	dl-Limonene	0.71	0.0102	
Hybrid	Immature	Monocopper oxidase-like protein SKU5 (Cc08_g12940)	dl-Limonene	0.66	0.0190	GO:0005507, GO:0055114, GO:0016491
Hybrid	Immature	Probable galacturonosyltransferase-like 3 (Cc11_g10580)	dl-Limonene	0.76	0.0040	
Inbred	Immature	Endoglucanase 11 (Cc05_g03130)	(E)-2-Decenal	0.66	0.0269	GO:0003824, GO:0005975
Hybrid	Immature	Endoglucanase 11 (Cc05_g03130)	dl-Limonene	0.71	0.0102	GO:0003824, GO:0005975
Hybrid	Immature	PREDICTED: At5g60140-like (B3 TF) (Cc11_g10560)	dl-Limonene	0.71	0.0100	GO:0003677
Inbred	Mature	Putative 14 kDa proline-rich protein (1) DC2.15 (Cc00_g00780)	(E)-2-Decenal	0.86	0.0007	GO:0006869
Hybrid	Mature	Putative 14 kDa proline-rich protein (1) DC2.15 (Cc00_g00780)	(E)-2-Decenal	0.74	0.0061	GO:0006869

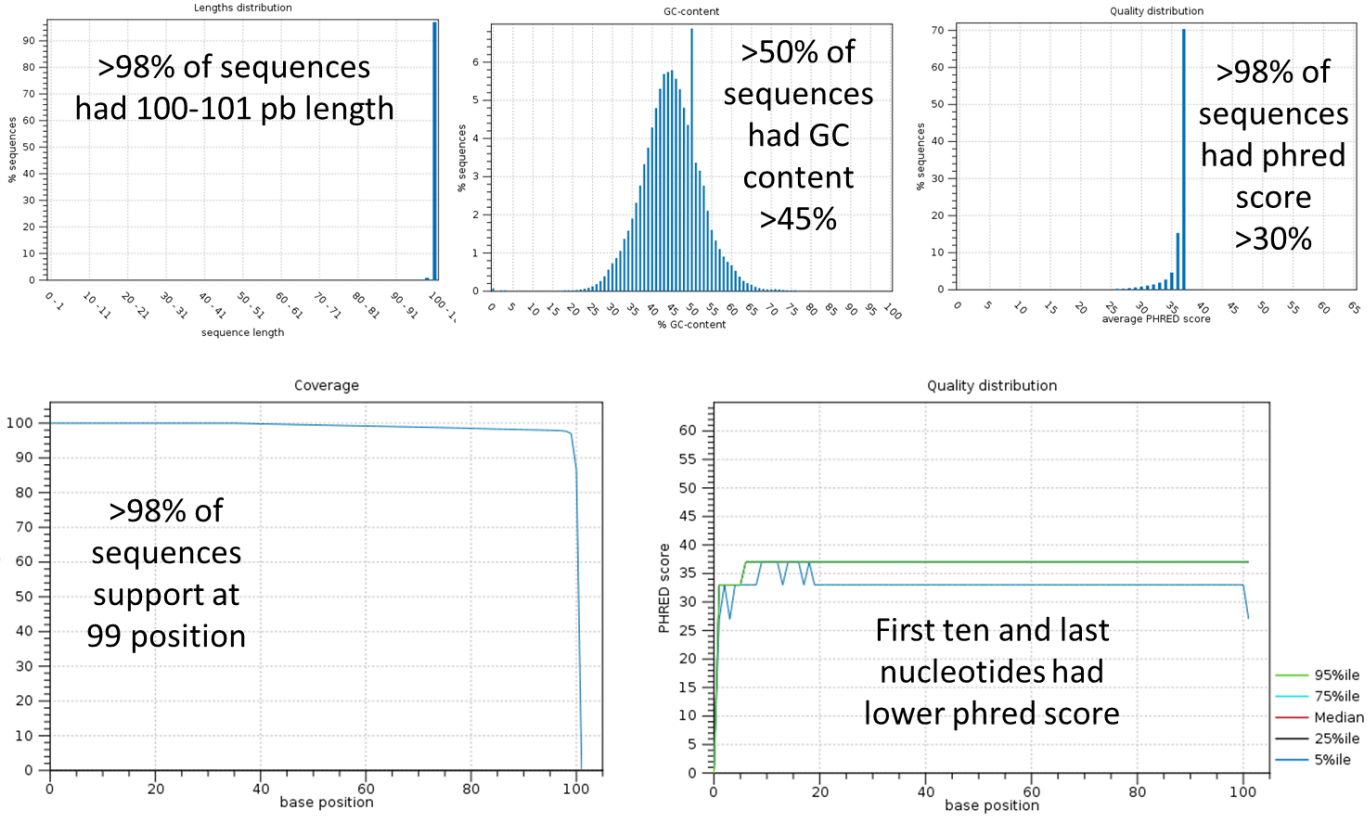
Appendix 12 continued…

Cultivar	Maturity stage	Candidate gene (Gene ID [134])	Volatile	Spearman p	*p*-value	Functional GO term(s) associated
Hybrid	Mature	Putative 14 kDa proline-rich protein (2) DC2.15 (Cc00_g00800)	(E,E)-2,4-Decadienal	0.74	0.0061	GO:0006869
Hybrid	Immature	Two-component response regulator-like APRR2 (Cc08_g02100)	dl-Limonene	0.72	0.0088	GO:0000160, GO:0000156, GO:0045449, GO:0003677
Hybrid	Immature	PREDICTED: uncharacterized protein LOC100260621	dl-Limonene	0.67	0.0171	
Hybrid	Immature	Putative pectinesterase 14 (Cc06_g04690)	dl-Limonene	0.66	0.0190	GO:0005618, GO:0042545, GO:0030599

APPENDIX 13

Summary of the quality report of the leaf sequences after trimming.

574,951,532 sequences from 23 samples



APPENDIX 14

Sequencing depth distribution for the leaf samples.

Tissue	Leaves											Overall
Cultivar	Inbred	Inbred	Inbred	Inbred	Hybrid	Hybrid	Hybrid	Hybrid	Inbred	Hybrid	Overall	
Treatment	R	C	R+T	T	R	C	R+T	T				
>0-0.5X	17,053	18,371	18,489	17,783	18,271	18,783	16,820	18,206	18,865	18,988	18,163	18,163
0.5-5X	5,357	4,385	3,875	4,801	4,381	3,689	5,498	4,367	4,673	4,558	4,558	4,558
5X-50X	424	290	261	321	277	256	413	298	325	319	318	318
50X-500X	30	14	16	16	14	14	26	13	19	13	18	18
>500X	1	0	0	0	0	0	1	0	0	0	0	0
Total higher than 0.5X	5,812	4,689	4,152	5,138	4,672	3,959	5,938	4,678	5,017	4,890	4,895	4,895
Percentage higher	25.4%	20.3%	18.3%	22.4%	20.4%	17.4%	26.1%	20.4%	21.0%	20.5%	21.2%	21.2%
Total expressed genes	22,865	23,060	22,641	22,921	22,943	22,742	22,758	22,884	23,882	23,878	23,057	23,057

APPENDIX 15

Quality report of the fragments and mapping of the leaf samples according to the treatments. The total, standard deviation (sd), and percent of each parameter are described for each treatment. Treatment number corresponds to Table 3.

Parameter	Treatment 1			Treatment 2			Treatment 3			Treatment 4			Treatment 5			Treatment 6			Treatment 7			Treatment 8			Average leaves		
	Total	sd	%	Total	sd	%	Total	sd	%	Total	sd	%	Total	sd	%	Total	sd	%	Total	sd	%	Total	sd	%	Total	sd	%
Counted fragments	3.0E+07	8.2E+06	96	2.2E+07	5.0E+05	91	2.0E+07	5.5E+05	96	2.3E+07	1.1E+06	92	2.3E+07	9.1E+05	96	2.0E+07	6.9E+05	93	3.1E+07	8.3E+06	97	2.2E+07	7.7E+05	94	2.4E+07	3.7E+06	94
unique fragments	2.9E+07	8.1E+06	93	2.1E+07	4.5E+05	88	2.0E+07	5.2E+05	93	2.3E+07	1.0E+06	89	2.2E+07	8.1E+05	93	1.9E+07	6.9E+05	89	3.0E+07	8.1E+06	94	2.1E+07	7.3E+05	91	2.3E+07	3.6E+06	91
non-specifically	8.5E+05	1.8E+05	3	7.8E+05	6.5E+04	3	6.4E+05	5.7E+04	3	7.7E+05	8.1E+04	3	7.4E+05	9.5E+04	3	6.6E+05	1.2E+04	3	9.2E+05	1.9E+05	3	7.2E+05	5.4E+04	3	7.6E+05	8.7E+04	3
Uncounted fragments	1.2E+06	2.8E+05	4	2.3E+06	8.9E+05	9	7.5E+05	1.9E+05	4	2.0E+06	6.3E+04	8	8.4E+05	1.3E+05	4	1.6E+06	6.7E+05	7	9.7E+05	2.7E+05	3	1.3E+06	3.5E+04	6	1.4E+06	5.1E+05	6
Total fragments	3.1E+07	8.2E+06	100	2.4E+07	1.0E+06	100	2.1E+07	6.0E+05	100	2.5E+07	1.1E+06	100	2.4E+07	1.0E+06	100	2.2E+07	1.3E+05	100	3.2E+07	8.6E+06	100	2.3E+07	7.4E+05	100	2.5E+07	3.6E+06	100
	Mapped	sd	%	Mapped	sd	%	Mapped	sd	%	Mapped	sd	%	Mapped	sd	%	Mapped	sd	%	Mapped	sd	%	Mapped	sd	%	Mapped	sd	%
Exon	1.9E+07	5.3E+06	63	1.4E+07	3.1E+05	61	1.3E+07	3.6E+05	62	1.4E+07	6.1E+05	62	1.4E+07	4.0E+05	61	1.3E+07	2.8E+05	63	1.9E+07	5.2E+06	61	1.4E+07	3.8E+05	62	1.5E+07	2.3E+06	62
Exon-exon	7.4E+06	2.2E+06	25	5.5E+06	9.2E+04	25	5.2E+06	8.4E+04	26	5.9E+06	3.3E+05	25	5.8E+06	2.3E+05	25	4.4E+06	3.5E+05	22	7.8E+06	2.3E+06	25	5.3E+06	1.4E+05	24	5.9E+06	1.1E+06	25
Total exon	2.6E+07	7.5E+06	88	1.9E+07	2.7E+05	86	1.8E+07	4.5E+05	88	2.0E+07	9.3E+05	87	2.0E+07	6.2E+05	86	1.7E+07	6.3E+05	85	2.7E+07	7.6E+06	87	1.9E+07	5.2E+05	86	2.1E+07	3.4E+06	87
Intron	8.5E+05	2.1E+05	3	7.0E+05	7.7E+04	3	6.1E+05	5.3E+04	3	8.1E+05	2.2E+04	3	9.7E+05	6.0E+04	4	6.0E+05	9.8E+04	3	1.3E+06	2.9E+05	4	7.6E+05	1.2E+05	3	8.2E+05	2.0E+05	3
Total gene	2.7E+07	7.7E+06	91	2.0E+07	3.5E+05	89	1.9E+07	4.5E+05	91	2.1E+07	9.5E+05	90	2.1E+07	6.4E+05	90	1.8E+07	7.3E+05	88	2.8E+07	7.9E+06	91	2.0E+07	5.9E+05	90	2.2E+07	3.5E+06	90
Intergenic	2.7E+06	5.3E+05	9	2.4E+06	1.8E+05	11	1.8E+06	2.1E+05	9	2.2E+06	2.7E+05	10	2.2E+06	2.7E+05	10	2.4E+06	6.2E+04	12	2.7E+06	4.5E+05	9	2.3E+06	2.3E+05	10	2.3E+06	2.5E+05	10
Total	3.0E+07	8.2E+06	100	2.2E+07	5.0E+05	100	2.0E+07	5.5E+05	100	2.3E+07	1.1E+06	100	2.3E+07	9.1E+05	100	2.0E+07	6.9E+05	100	3.1E+07	8.3E+06	100	2.2E+07	7.7E+05	100	2.4E+07	3.7E+06	100

APPENDIX 16

Total 136 differentially expressed genes (DEGs) found when compared the inbred vs hybrid with higher sequence depth (>0.5X) and without unknown or uncharacterized annotated descriptions. Gene ID and annotation are displayed according to the reference genome [134]. Average sequencing depth was made by averaging all samples and treatments as described by Dugas, et al. [136]. The GO term number was presented for known genes using AgriGO 2.0 [188]. Fold change (FC) higher than 2.0 represents up-regulated expression (increased in the hybrid), while lower than 2.0 represents down-regulated expression (increased in the inbred). Statistical significance using Bonferroni and FDR are shown. The data is ordered according to decreasing fold change.

Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
Cc08_g16300	Putative GlutaRedoxin family protein	1.800	GO:0045454	11.384	0.000	0.000
Cc00_g20080	ABC transporter G family member 11	4.592	GO:0016020	8.970	0.000	0.000
Cc05_g05180	Putative B-box type zinc finger family protein	0.914	GO:0008270	6.855	0.000	0.000
Cc02_g05510	unknown protein; Has 39 Blast hits to 39 proteins in 15 species	2.926	N/A	6.385	0.000	0.000
Cc04_g10750	BTB/POZ and TAZ domain-containing protein 2	1.486	GO:0005634	5.137	0.000	0.000
Cc06_g18960	Putative N-acylneuraminate-9-phosphatase	6.767	GO:0008967	4.850	0.000	0.000
Cc02_g36770	Ferric reduction oxidase 7, chloroplastic	6.880	GO:0050660	4.545	0.000	0.000
Cc08_g08370	Caffeic acid 3-O-methyltransferase	0.547	GO:0046983	4.151	0.000	0.000
Cc01_g20020	Histone H2AX	0.575	GO:0005634	4.089	0.000	0.000
Cc06_g13690	Putative Zinc finger protein CONSTANS-LIKE 4	1.029	GO:0008270	3.971	0.000	0.000
Cc05_g12000	Protein SPA1-RELATED 4	0.598	N/A	3.664	0.000	0.000
Cc05_g02850	Putative Glucan endo-1,3-beta-glucosidase 3	0.530	GO:0003824	3.602	0.000	0.000
Cc02_g39100	Hydroquinone glucosyltransferase	4.202	GO:0008152	3.535	0.000	0.000
Cc07_g11750	Undecaprenyl diphosphate synthase, putative	1.563	N/A	3.523	0.000	0.000
Cc07_g02190	Putative L-aspartate oxidase	2.068	GO:0005737	3.322	0.000	0.000
Cc10_g11980	Chlorophyllide a oxygenase, chloroplastic	0.775	GO:0005506	3.220	0.000	0.000
Cc02_g03440	Putative Post-GPI attachment to proteins factor 3	0.663	N/A	3.159	0.000	0.000

Appendix 16 continued…

Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
Cc07_g01090	Putative Glycerophosphodiester phosphodiesterase gde1	2.746	GO:0008889	3.144	0.000	0.000
Cc04_g09270	Purple acid phosphatase 17	2.182	GO:0016787	3.094	0.004	0.000
Cc10_g02310	ATP-dependent protease La (LON) domain protein	1.973	GO:0004176	3.026	0.000	0.000
Cc11_g05930	Putative Probable transmembrane ascorbate ferrireductase 3	1.840	GO:0016021	3.015	0.000	0.000
Cc11_g13920	Putative Transcription factor MYB1R1	0.769	GO:0003677	2.921	0.000	0.000
Cc02_g38130	Glycerol-3-phosphate acyltransferase 4	0.601	GO:0008415	2.920	0.000	0.000
Cc10_g13610	Putative Probable calcium-binding protein CML44	2.397	GO:0005509	2.857	0.000	0.000
Cc07_g10470	Putative Probable salt tolerance-like protein At1g78600	1.072	GO:0008270	2.836	0.000	0.000
Cc06_g08450	SPX domain-containing protein 2	2.340	N/A	2.818	0.000	0.000
Cc02_g00820	Putative Two-component response regulator-like APRR5	3.399	GO:0000160	2.811	0.000	0.000
Cc02_g19370	unknown protein; BEST Arabidopsis thaliana protein match is	2.039	N/A	2.793	0.000	0.000
Cc07_g02110	NAD(P)-binding Rossmann-fold superfamily protein	1.148	GO:0003824	2.791	0.000	0.000
Cc00_g18980	Putative Probable receptor-like protein kinase At5g39020	0.660	GO:0004674	2.777	0.000	0.000
Cc05_g13590	fatty acid desaturase A	2.250	N/A	2.754	0.000	0.000
Cc02_g09170	Putative Trans-2,3-enoyl-CoA reductase	0.769	GO:0005737	2.749	0.000	0.000
Cc03_g15890	F-box/ankyrin repeat protein SKIP35	1.624	N/A	2.660	0.000	0.000
Cc05_g07810	GlutaRedoxin-C9	0.909	GO:0045454	2.657	0.000	0.000
Cc02_g24440	SPX domain-containing protein 3	0.841	N/A	2.634	0.000	0.000
Cc02_g04050	DNA photolyases;DNA photolyases	1.042	GO:0003913	2.624	0.000	0.000
Cc10_g00330	plastid movement impaired1	0.716	N/A	2.549	0.000	0.000
Cc00_g16400	(+)-neomenthol dehydrogenase	1.517	GO:0055114	2.465	0.000	0.000
Cc03_g06030	Protein PROTON GRADIENT REGULATION 5, chloroplastic	5.380	N/A	2.405	0.000	0.000
Cc10_g01850	ATP sulfurylase 2	0.518	GO:0004781	2.403	0.000	0.000
Cc00_g05460	Pathogenesis-related protein 5	1.065	N/A	2.382	0.001	0.000

Appendix 16 continued…

Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
Cc00_g26120	SC35-like splicing factor 33	0.519	GO:0003676	2.351	0.000	0.000
Cc03_g00700	Putative Ca2+-binding protein 1	2.435	GO:0005509	2.336	0.000	0.000
Cc05_g07890	8-hydroxyquercetin 8-O-methyltransferase	2.474	GO:0046983	2.336	0.000	0.000
Cc00_g24160	Putative dihydrodipicolinate reductase 3, chloroplastic	0.519	GO:0005737	2.323	0.000	0.000
Cc10_g02730	RAD-like 1	0.853	GO:0003677	2.320	0.010	0.000
Cc06_g07500	Probable protein phosphatase 2C 27	1.137	GO:0003824	2.298	0.000	0.000
Cc04_g16990	Putative Pectinesterase	0.671	GO:0004857	2.294	0.000	0.000
Cc01_g09970	Putative Bifunctional polymyxin resistance protein ArnA	4.006	GO:0003824	2.287	0.000	0.000
Cc02_g20030	Putative phytosulfokine 4 precursor	0.682	GO:0005576	2.282	0.000	0.000
Cc07_g08970	Putative Rhomboid-related intramembrane serine protease family protein	0.605	GO:0004252	2.262	0.000	0.000
Cc11_g15740	Inorganic phosphate transporter 2-1, chloroplastic	1.528	GO:0006817	2.256	0.000	0.000
Cc02_g26830	Putative Uncharacterized GPI-anchored protein At1g27950	0.885	GO:0008289	2.253	0.000	0.000
Cc06_g15890	Actin-97	1.632	GO:0005515	2.249	0.000	0.000
Cc02_g07890	Actin	4.722	GO:0005524	2.249	0.000	0.000
Cc02_g39110	Putative Dihydroflavonol-4-reductase	0.829	GO:0003824	2.248	0.000	0.000
Cc00_g28800	Putative Isoflavone 7-O-methyltransferase	1.758	GO:0046983	2.241	0.001	0.000
Cc05_g09930	Chlorophyll a-b binding protein 8, chloroplastic	18.224	GO:0016020	2.227	0.002	0.000
Cc07_g00260	Chlorophyll a-b binding protein 13, chloroplastic	3.158	GO:0016020	2.206	0.000	0.000
Cc10_g08170	Glutamyl-tRNA reductase 1, chloroplastic	2.887	GO:0005737	2.199	0.002	0.000
Cc06_g04970	Sulfite exporter TauE/SafE family protein	0.562	GO:0016021	2.179	0.000	0.000
Cc11_g14720	Carbonic anhydrase, chloroplastic	4.534	GO:0015976	2.167	0.000	0.000
Cc04_g08250	Putative Dof zinc finger protein DOF5.4	0.534	GO:0045449	2.163	0.000	0.000
Cc03_g15090	UDP-glycosyltransferase 74F2	1.149	GO:0008152	2.161	0.000	0.000
Cc05_g12720	Chlorophyll a-b binding protein 13, chloroplastic	8.688	GO:0016020	2.157	0.000	0.000

Appendix 16 continued…

Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
Cc03_g02680	unknown protein; BEST Arabidopsis thaliana protein match is	2.692	N/A	2.135	0.000	0.000
Cc11_g13590	Nitrate transporter 1.7	0.571	GO:0006857	2.118	0.000	0.000
Cc06_g11160	Serine acetyltransferase 5	0.687	GO:0005737	2.103	0.000	0.000
Cc02_g37510	(+)-neomenthol dehydrogenase	15.768	GO:0055114	2.102	0.001	0.000
Cc02_g24580	Probable plastid-lipid-associated protein 11, chloroplastic	0.938	GO:0009507	2.082	0.007	0.000
Cc02_g06400	Omega-3 fatty acid desaturase, chloroplastic	3.968	GO:0006629	2.080	0.000	0.000
Cc02_g39990	Putative Protein LHY	1.531	GO:0003677	2.071	0.000	0.000
Cc01_g17770	Photosystem II reaction center PSB28 protein, chloroplastic	1.372	GO:0016020	2.067	0.000	0.000
Cc11_g10250	Probable nitrite transporter At1g68570	2.600	GO:0006857	2.060	0.007	0.000
Cc08_g00770	SPX domain-containing protein 2	1.205	N/A	2.030	0.000	0.000
Cc01_g10520	Putative Heavy metal transport/detoxification superfamily protein	0.527	GO:0046872	2.028	0.000	0.000
Cc00_g05390	Putative Cation transport regulator-like protein 2	1.072	N/A	2.024	0.000	0.000
Cc07_g13520	NTA15 protein	0.604	N/A	2.020	0.000	0.000
Cc01_g11710	Putative DDB1- and CUL4-associated factor 8	1.175	N/A	2.013	0.000	0.000
Cc06_g09790	Putative Pentatricopeptide repeat-containing protein At5g21222	0.736	N/A	2.004	0.000	0.000
Cc03_g03270	Putative unknown protein; BEST Arabidopsis thaliana protein match is	1.604	N/A	-2.029	0.000	0.000
Cc08_g15910	Putative Probable WRKY transcription factor 41	0.566	GO:0003700	-2.076	0.000	0.000
Cc01_g21030	Putative Epoxide hydrolase 2	0.512	GO:0003824	-2.087	0.000	0.000
Cc06_g06600	BEL1-like homeodomain protein 1	3.032	GO:0003700	-2.115	0.000	0.000
Cc11_g05610	NAD(P)H-dependent 6'-deoxychalcone synthase	0.730	GO:0055114	-2.126	0.008	0.000
Cc09_g06130	Putative Uncharacterized protein yqjG	0.833	N/A	-2.143	0.001	0.000
Cc07_g09950	P-loop containing nucleoside triphosphate hydrolases superfamily protein	0.569	GO:0017111	-2.146	0.000	0.000
Cc06_g22450	Putative Cytochrome P450 83B1	0.746	GO:0004497	-2.173	0.000	0.000

Appendix 16 continued…

Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
Cc02_g28890	Serine carboxypeptidase-like 40	1.259	GO:0006508	-2.188	0.000	0.000
Cc02_g25540	Protein of unknown function (DUF581)	1.454	N/A	-2.198	0.000	0.000
Cc09_g10500	Lysine histidine transporter-like 8	0.710	N/A	-2.214	0.000	0.000
Cc00_g30530	Putative Cytochrome P450 83B1	0.816	GO:0004497	-2.263	0.000	0.000
Cc05_g08830	Secologanin synthase	0.540	GO:0004497	-2.285	0.000	0.000
Cc03_g02380	Inactive beta-amylase 9	1.250	GO:0003824	-2.314	0.000	0.000
Cc06_g23270	Putative F-box/kelch-repeat protein At1g15670	1.550	N/A	-2.335	0.000	0.000
Cc02_g20290	Pathogenesis-related protein 1B	24.425	GO:0005576	-2.345	0.002	0.000
Cc04_g17250	AAA-ATPase 1	0.534	GO:0017111	-2.345	0.000	0.000
Cc10_g04940	Putative Transcription factor MYB44	1.047	GO:0045449	-2.368	0.000	0.000
Cc02_g12240	Major facilitator superfamily protein	1.893	GO:0016020	-2.388	0.000	0.000
Cc04_g01330	Nodulation-signaling pathway 1 protein	1.121	N/A	-2.425	0.000	0.000
Cc04_g11510	Tropinone reductase homolog At1g07440	0.889	GO:0005488	-2.434	0.006	0.000
Cc07_g08550	Probable non-specific lipid-transfer protein AKCS9	2.188	GO:0006869	-2.525	0.000	0.000
Cc07_g11550	Basic endochitinase A	1.062	GO:0004568	-2.534	0.000	0.000
Cc06_g18240	Glucan endo-1,3-beta-glucosidase, acidic isoform PR-Q'	6.225	GO:0003824	-2.564	0.000	0.000
Cc06_g13940	EG45-like domain containing protein	4.617	N/A	-2.654	0.009	0.000
Cc00_g14270	NAC domain containing protein 36	1.043	GO:0045449	-2.663	0.000	0.000
Cc08_g09450	Putative UPF0481 protein At3g02645	0.538	N/A	-2.707	0.000	0.000
Cc03_g00450	galactinol synthase 2	2.128	N/A	-2.707	0.004	0.000
Cc06_g07210	Putative Lysosomal beta glucosidase	0.685	GO:0005975	-2.796	0.000	0.000
Cc11_g10010	Putative unknown protein; BEST Arabidopsis thaliana protein match is	0.547	N/A	-2.797	0.001	0.000
Cc00_g13890	Double WRKY type transfactor	0.570	GO:0003700	-2.817	0.000	0.000
Cc07_g07790	Uncharacterized protein At5g65660	0.726	N/A	-2.820	0.000	0.000
Cc10_g03880	Protein of unknown function, DUF617	0.570	N/A	-2.851	0.000	0.000
Cc04_g05080	Probable WRKY transcription factor 40	0.516	GO:0003700	-2.885	0.000	0.000
Cc07_g13290	Glutamine synthetase cytosolic isozyme 2	1.947	GO:0004356	-2.941	0.001	0.000
Cc03_g03550	Sugar transport protein 13	1.514	GO:0022891	-2.947	0.000	0.000

Appendix 16 continued…

Gene ID [134]	Annotation description	Average sequencing depth	GO term	Fold change	Bonferroni	FDR *p*-value correction
Cc00_g14300	Basic endochitinase	12.307	GO:0004568	-2.982	0.000	0.000
Cc10_g11110	Putative Vinorine synthase	1.291	N/A	-3.014	0.000	0.000
Cc11_g07780	Pathogenesis-related protein R major form	5.451	N/A	-3.043	0.003	0.000
Cc04_g03730	Pleiotropic drug resistance protein 1	2.748	GO:0016020	-3.125	0.000	0.000
Cc00_g34980	UDP-glucose flavonoid 3-O-glucosyltransferase 6	0.805	GO:0008152	-3.171	0.000	0.000
Cc00_g34570	CYP71AT2v1	1.318	GO:0004497	-3.283	0.000	0.000
Cc09_g01430	WRKY transcription factor 6	0.586	GO:0003700	-3.287	0.000	0.000
Cc07_g13880	Probable 1-deoxy-D-xylulose-5-phosphate synthase 2, chloroplastic	0.903	GO:0003824	-3.341	0.000	0.000
Cc11_g00410	Putative Acidic endochitinase	40.277	GO:0003824	-3.358	0.000	0.000
Cc00_g11200	Probable LRR receptor-like serine/threonine-protein kinase At4g08850	0.791	GO:0006468	-3.577	0.000	0.000
Cc06_g18280	Glucan endo-1,3-beta-glucosidase, acidic isoform PR-Q'	7.356	GO:0003824	-3.645	0.000	0.000
Cc00_g33180	Endochitinase B	8.574	GO:0004568	-3.731	0.000	0.000
Cc08_g14230	Thaumatin-like protein	1.818	N/A	-3.775	0.000	0.000
Cc10_g10180	serine-type endopeptidase inhibitors	0.903	GO:0004867	-3.778	0.004	0.000
Cc06_g09330	Predicted protein	1.022	N/A	-4.206	0.001	0.000
Cc03_g09010	Putative CYC02 protein	5.003	N/A	-5.160	0.000	0.000
Cc08_g01530	Beta-amylase 3, chloroplastic	0.931	GO:0003824	-7.817	0.000	0.000
Cc05_g12500	NAC domain-containing protein 72	0.825	GO:0045449	-10.490	0.000	0.000
Cc11_g12740	NAC domain-containing protein 29	2.419	GO:0045449	-10.582	0.000	0.000
Cc04_g10600	Putative Cytochrome P450 82A3	0.602	GO:0004497	-306.387	0.000	0.000

APPENDIX 17

Stepwise regression's estimated DEGs, significance, trait prediction, and GO term classification. The treatment represents the condition by which the trait associated is expected to vary. The regulation represents if the gene expression was up-regulated (increased with the trait value increase), or was down-regulated (decreased with increase of the trait value). The estimate represents how much the gene expression effect increased (positive value) or decreased (negative value) the value of the trait. GO terms with several pathways, are referred as "Several", rather than specific GO term code.

Cultivar	Treat.	Trait	Regulation	Annotation (Gene ID [134])	Estimate	Std Error	Prob>|*t*|	GO term
Inbred	R	TH	Down	Putative BON1-associated protein 2 (Cc00_g30940)	-0.02683	9.42E-06	0.0002	Several
Inbred	R	OC	Down	Putative uncharacterized protein (Cc00_g03090)	0.000973	5.64E-05	0.0033	Several
Inbred	R	TH	Down	Beta-fructofuranosidase, insoluble isoenzyme 1 (Cc06_g06090)	-0.00288	3.23E-06	0.0007	GO:0005975
Hybrid	R	TL	Down	40S ribosomal protein S27-1 (Cc11_g14900)	-0.00138	2.32E-07	0.0001	GO:0003735
Hybrid	R	TL	Down	40S ribosomal protein S27-1 (Cc11_g14900)	-0.00138	2.32E-07	0.0001	GO:0003735
Inbred	R	RI	Down	Putative Cytochrome P450 87A3 (Cc02_g08960)	1.37E-05	3.11E-10	<.0001	GO:0004497
Inbred	R	RI	Down	Cytochrome P450 87A3 (Cc00_g06850)	-0.00056	4.16E-10	<.0001	GO:0004497
Hybrid	R	RI	Down	Putative Calcium-dependent lipid-binding (CaLB domain) family protein (Cc00_g10550)	-6.60E-05	1.55E-09	<.0001	Several
Hybrid	R	TH	Up	Putative Disease resistance protein RGA2 (Cc00_g26280)	0.143057	3.45E-06	<.0001	GO:0006952
Hybrid	R	TL	Down	Sucrose synthase 2 (Cc04_g08870)	4.90E-05	1.12E-07	0.0015	GO:0009058
Hybrid	R	TL	Down	Hypothetical protein (Cc00_g07620)	0.00029	1.67E-07	0.0004	Several

Appendix 17 continued…

Cultivar	Treat.	Trait	Regulation	Annotation (Gene ID [134])	Estimate	Std Error	Prob>|*t*|	GO term
Hybrid	R	RS	Up	Putative uncharacterized protein (Cc00_g03090)	0.000153	3.76E-06	<.0001	Several
Hybrid	R	RI	Up	Putative Aldo-keto reductase yakc (Cc02_g36130)	-0.00324	4.93E-06	<.0001	GO:0055114
Hybrid	R	TL	Up	Hypothetical protein (Cc00_g07620)	-0.0054	1.45E-05	<.0001	Several
Hybrid	R	RI	Up	Hypothetical protein (Cc00_g07620)	-2.90E-05	5.55E-07	<.0001	Several
Hybrid	R	RE	Down	Putative Monoglyceride lipase (Cc00_g15700)	0.000576	1.71E-08	<.0001	Several
Hybrid	R	OC	Up	Putative Disease resistance protein RGA2 (Cc00_g26280)	-0.00068	2.09E-09	<.0001	GO:0006952
Hybrid	R	TL	Up	Putative ABC transporter A family member 12 (Cc00_g22610)	0.020609	8.17E-06	<.0001	Several
Hybrid	R	OC	Up	Putative ABC transporter A family member 12 (Cc00_g22610)	0.002559	1.28E-09	<.0001	Several
Hybrid	R	RI	Up	Putative ABC transporter A family member 12 (Cc00_g22610)	-0.00024	5.66E-07	<.0001	Several
Hybrid	R	RE	Down	UDP-glycosyltransferase 85A3 (Cc00_g30680)	-0.00015	6.25E-08	0.0003	GO:0008152
Hybrid	R	OC	Up	Putative Small subunit processome component 20 homolog (Cc00_g24200)	0.002483	2.76E-09	<.0001	GO:0005488
Inbred	R	RE	Down	Putative BON1-associated protein 2 (Cc00_g30940)	0.002078	3.99E-05	0.0004	Several
Hybrid	R	RI	Down	Small nuclear ribonucleoprotein family protein (Cc02_g33830)	-0.00019	7.41E-09	<.0001	Several
Inbred	R	RI	Down	Major allergen Pru ar 1 (Cc03_g01290)	0.000377	1.07E-09	<.0001	GO:0006952
Hybrid	R	TH	Down	Putative unknown protein; INVOLVED IN (Cc08_g08670)	-0.02261	0.001706	0.0009	Several
Hybrid	R	RE	Down	RING/FYVE/PHD zinc finger superfamily protein (Cc00_g23100)	-5.90E-05	1.46E-08	0.0002	GO:0008270
Hybrid	R	TL	Down	Hypothetical protein (Cc00_g07620)	-0.00898	1.25E-06	<.0001	Several
Hybrid	R	TH	Down	Basic endochitinase (Cc00_g14300)	0.003663	0.000259	0.0008	GO:0004568
Hybrid	R	TL	Down	Mannan endo-1,4-beta-mannosidase 5 (Cc00_g30210)	0.016176	0.000888	<.0001	GO:0003824

Appendix 17 continued…

Cultivar	Treat.	Trait	Regulation	Annotation (Gene ID [134])	Estimate	Std Error	Prob>|*t*|	GO term
Inbred	R	TH	Down	Mannan endo-1,4-beta-mannosidase 5 (Cc00_g30210)	0.026252	1.26E-05	0.0003	GO:0003824
Hybrid	R	TL	Down	Hypothetical protein (Cc00_g07620)	0.018012	5.33E-06	0.0002	Several
Hybrid	R	RS	Up	Putative pentatricopeptide repeat-containing protein At1g12700, mitochondrial (Cc00_g04090)	0.000161	2.21E-06	<.0001	Several
Hybrid	R	RE	Down	Pathogenesis-related genes transcriptional activator PTI5 (Cc02_g14240)	8.16E-05	9.26E-08	0.0007	GO:0003700
Hybrid	R	RS	Down	Arginine decarboxylase (Cc02_g16860)	2.05E-07	8.77E-09	0.0018	GO:0008792
Hybrid	R	TL	Up	Oligopeptide transporter 4 (Cc02_g20620)	-0.01601	8.25E-06	<.0001	GO:0055085
Hybrid	R	RI	Up	Putative Probable 1,4-dihydroxy-2-naphthoate octaprenyltransferase (Cc02_g21250)	-4.10E-06	2.77E-07	0.0007	GO:0016021
Hybrid	R	TL	Down	Pectinesterase 2 (Cc02_g24210)	-0.0113	0.000249	<.0001	GO:0005618
Hybrid	R	RS	Down	Putative Receptor-like protein 12 (Cc11_g00880)	-2.30E-05	6.42E-08	<.0001	GO:0005515
Hybrid	R	RE	Down	Putative Receptor-like protein 12 (Cc01_g08260)	-2.00E-05	7.18E-09	0.0002	GO:0005515
Hybrid	R	TL	Down	Putative Basic 7S globulin (Cc00_g13300)	0.038096	0.000199	<.0001	GO:0006508
Hybrid	R	RE	Up	Putative Myosin-J heavy chain (Cc01_g17540)	-4.80E-05	1.27E-06	<.0001	GO:0016459
Hybrid	R	OC	Up	Pentatricopeptide repeat-containing protein At5g48730, chloroplastic (Cc01_g16930)	0.000293	2.45E-09	<.0001	Several
Inbred	R	RS	Down	Tyrosine aminotransferase (Cc05_g07600)	0.000413	6.64E-05	0.0008	GO:0016847
Inbred	R	OC	Down	Hypothetical protein (Cc00_g07620)	0.000666	3.13E-06	<.0001	Several
Inbred	R	RI	Down	Hypothetical protein (Cc00_g07620)	-1.90E-05	2.36E-10	<.0001	Several
Hybrid	R	TH	Up	Hypothetical protein (Cc00_g07620)	0.000425	1.83E-06	<.0001	Several
Hybrid	R	TL	Up	Protein of unknown function (DUF707) (Cc09_g03510)	-3.90E-05	3.51E-06	0.0078	Several
Inbred	R	TH	Down	Pathogenesis-related protein R major form (Cc11_g07800)	-0.00012	1.54E-08	<.0001	Several
Hybrid	R	RI	Down	Ammonium transporter 3 member 1 (Cc07_g19370)	-1.90E-05	5.38E-09	0.0002	GO:0016020

Appendix 17 continued…

Cultivar	Treat.	Trait	Regulation	Annotation (Gene ID [134])	Estimate	Std Error	Prob>|*t*|	GO term
Hybrid	R	RE	Down	Nuclear transport factor 2 (Cc07_g00920)	0.000392	2.15E-08	<.0001	GO:0006810
Hybrid	R	RE	Up	Putative unknown protein; Has 50 Blast hits to 45 proteins in 14 species (Cc03_g11440)	-0.0016	9.79E-06	<.0001	Several
Hybrid	R	TH	Down	Acidic endochitinase SE2 (Cc05_g00810)	-0.00148	0.000144	0.002	GO:0003824
Hybrid	R	RE	Down	40S ribosomal protein S10 (Cc02_g02340)	-4.70E-05	4.82E-09	<.0001	Several
Hybrid	R	RE	Down	Carboxylesterase 1 (Cc02_g03630)	-0.00021	2.72E-08	<.0001	GO:0016787
Hybrid	R	TL	Up	Sodium/hydrogen exchanger 7 (Cc11_g05270)	-0.00281	2.06E-05	<.0001	GO:0006812
Hybrid	R	RI	Down	Putative LRR receptor-like serine/threonine-protein kinase FLS2 (Cc11_g06630)	0.001	5.22E-08	<.0001	GO:0005515
Hybrid	R	TH	Down	Putative Leucine-rich repeat receptor-like protein kinase PXL1 (Cc11_g06760)	0.011413	0.001395	0.0038	GO:0005515
Hybrid	R	RI	Down	Putative Leucine-rich repeat receptor-like protein kinase TDR (Cc00_g31890)	-9.40E-05	7.57E-09	<.0001	GO:0005515
Hybrid	R	RS	Up	ABC transporter C family member 1 (Cc09_g08460)	0.000214	6.31E-07	<.0001	GO:0016021
Inbred	R	RS	Down	1-aminocyclopropane-1-carboxylate oxidase 3 (Cc05_g02900)	9.76E-06	1.06E-06	<.0001	GO:0016491
Inbred	R	RE	Down	Putative Hyoscyamine 6-dioxygenase (Cc05_g10390)	0.000333	3.57E-06	0.0001	GO:0016491
Inbred	R	RS	Down	Probable glutathione S-transferase parC (Cc07_g20060)	0.000152	1.71E-05	0.0001	Several
Inbred	R	OC	Down	Probable glutathione S-transferase (Cc05_g10080)	0.005001	4.32E-05	<.0001	Several
Hybrid	R	TL	Down	Tubulin beta-2 chain (Cc04_g06320)	-0.00089	1.39E-07	<.0001	GO:0051258
Hybrid	R	TH	Down	Purple acid phosphatase 17 (Cc04_g09270)	0.000619	3.86E-05	0.0005	GO:0016787
Inbred	R	RI	Down	phosphate transporter 3;1 (Cc04_g09390)	0.00027	2.00E-09	<.0001	GO:0016020
Hybrid	R	OC	Up	Putative Glycogen synthase (Cc09_g09040)	1.02E-05	6.09E-10	<.0001	GO:0009058

Appendix 17 continued…

Cultivar	Treat.	Trait	Regulation	Annotation (Gene ID [134])	Estimate	Std Error	Prob>|*t*|	GO term
Hybrid	R	OC	Up	Putative Glycogen synthase (Cc09_g09040)	1.02E-05	6.09E-10	<.0001	GO:0009058
Hybrid	R	RE	Up	Putative Glycogen synthase (Cc09_g09040)	5.21E-05	8.49E-07	<.0001	GO:0009058
Hybrid	R	RE	Up	Putative Glycosyl transferase, family 8 (Cc06_g00020)	-0.0018	2.92E-05	<.0001	Several
Hybrid	R	OC	Up	Putative uncharacterized protein Sb01g036200 (Cc06_g00350)	-0.00058	7.16E-09	<.0001	GO:0016020
Inbred	R	OC	Down	Probable anion transporter 3, chloroplastic (Cc06_g01670)	-6.20E-05	5.24E-06	0.0071	GO:0055085
Hybrid	R	RS	Down	PAR1 protein (Cc06_g01870)	5.81E-05	1.27E-07	<.0001	Several
Inbred	R	RI	Down	PAR1 protein (Cc06_g01870)	-7.10E-05	2.62E-10	<.0001	Several
Inbred	R	TH	Down	PAR1 protein (Cc06_g01870)	-0.00057	4.84E-07	0.0005	Several
Hybrid	R	RI	Down	Probable glutathione S-transferase (Cc05_g10080)	-7.70E-07	1.90E-09	0.0016	Several
Hybrid	R	TL	Down	V-type proton ATPase 16 kDa proteolipid subunit (Cc10_g02030)	0.001771	1.04E-07	<.0001	GO:0015986
Hybrid	R	TL	Down	V-type proton ATPase 16 kDa proteolipid subunit (Cc10_g02030)	0.001771	1.04E-07	<.0001	GO:0015986
Inbred	R	RE	Down	Putative Anthocyanidin 3-O-glucosyltransferase 5 (Cc10_g03270)	-0.00048	1.54E-05	0.001	GO:0008152
Inbred	R	TL	Down	Probable calcium-binding protein CML41 (Cc10_g03450)	0.037592	0.000955	<.0001	GO:0005509
Inbred	R	OC	Down	Probable calcium-binding protein CML41 (Cc10_g03450)	0.004437	0.000127	0.0008	GO:0005509
Inbred	R	RE	Down	Probable calcium-binding protein CML41 (Cc10_g03450)	0.001659	6.15E-05	0.0014	GO:0005509
Hybrid	R	TL	Up	Putative Probable serine/threonine-protein kinase DDB_G0276461 (Cc02_g27140)	-0.00557	0.000027	<.0001	GO:0004674
Hybrid	R	RE	Up	Putative Probable serine/threonine-protein kinase DDB_G0276461 (Cc02_g27140)	0.000613	7.23E-06	<.0001	GO:0004674
Hybrid	R	RE	Up	Putative uncharacterized protein (Cc00_g03090)	2.19E-05	1.94E-06	0.0015	Several
Inbred	R	RS	Down	Putative Vetispiradiene synthase 1 (Cc08_g06960)	-0.00029	2.85E-05	<.0001	GO:0016829

Appendix 17 continued…

Cultivar	Treat.	Trait	Regulation	Annotation (Gene ID [134])	Estimate	Std Error	Prob>|*t*|	GO term
Hybrid	R	OC	Up	Remorin family protein (Cc11_g03240)	0.000149	1.71E-09	<.0001	Several
Hybrid	R	OC	Up	Remorin family protein (Cc11_g03240)	0.000149	1.71E-09	<.0001	Several
Hybrid	R	TH	Up	Remorin family protein (Cc11_g03240)	-0.00131	6.88E-07	<.0001	Several
Hybrid	R	TL	Down	Beta-D-xylosidase 1 (Cc03_g01070)	-0.00091	3.22E-05	<.0001	GO:0005975
Hybrid	R	OC	Up	Hypothetical protein (Cc00_g07620)	-4.00E-06	2.60E-09	0.0004	Several
Hybrid	R	RE	Up	Hypothetical protein (Cc00_g07620)	-0.00187	5.35E-06	<.0001	Several
Hybrid	R	OC	Up	Putative Non-lysosomal glucosylceramidase (Cc03_g07570)	-0.00172	2.26E-09	<.0001	Several
Hybrid	R	RS	Up	Putative Non-lysosomal glucosylceramidase (Cc03_g07570)	-4.50E-06	5.12E-07	0.0032	Several
Inbred	R	TH	Down	Putative Reticuline oxidase-like protein (Cc03_g15300)	-0.00012	1.21E-06	0.0062	GO:0050660
Inbred	R	RI	Down	1-aminocyclopropane-1-carboxylate synthase (Cc02_g38530)	9.02E-06	2.21E-09	0.0002	GO:0003824
Hybrid	R	TH	Down	Chitinase 2 (Cc03_g12410)	-0.00408	0.000318	0.001	GO:0003824
Hybrid	R	RI	Up	Putative Pentatricopeptide repeat-containing protein At3g48810 (Cc01_g05040)	5.49E-05	1.56E-06	<.0001	Several
Hybrid	R	RI	Up	Protein of unknown function (DUF581) (Cc04_g00800)	1.33E-05	8.06E-08	<.0001	Several
Hybrid	R	RS	Down	Flavoprotein WrbA (Cc02_g11960)	7.58E-06	2.11E-07	0.0008	GO:0010181
Inbred	R	RE	Down	Putative Probable E3 ubiquitin-protein ligase ARI7 (Cc02_g10790)	-0.00331	2.04E-05	<.0001	GO:0008270
Hybrid	R	RE	Down	Phosphoenolpyruvate carboxylase kinase 2 (Cc02_g08980)	-5.80E-07	7.09E-09	0.0077	GO:0004674
Hybrid	R	RS	Down	Thaumatin-like protein (Cc08_g14230)	2.01E-05	5.09E-08	<.0001	Several
Inbred	R	RE	Down	Putative Acidic mammalian chitinase (Cc06_g15430)	0.001454	3.23E-05	0.0005	GO:0003824
Inbred	R	OC	Down	Putative Stellacyanin (Cc06_g08240)	-0.00099	3.47E-06	<.0001	GO:0005507
Inbred	R	RI	Down	Putative Stellacyanin (Cc06_g08240)	2.71E-05	1.30E-10	<.0001	GO:0005507
Inbred	R	TH	Down	serine-type endopeptidase inhibitors (Cc10_g10180)	0.001532	5.38E-07	0.0002	GO:0004867

Appendix 17 continued…

Cultivar	Treat.	Trait	Regulation	Annotation (Gene ID [134])	Estimate	Std Error	Prob>|*t*|	GO term
Hybrid	R	TL	Down	Putative Strictosidine synthase 1 (Cc05_g04400)	0.002717	2.25E-06	0.0005	GO:0016844
Inbred	R	TH	Down	1-aminocyclopropane-1-carboxylate synthase (Cc02_g38530)	-0.00074	3.22E-06	0.0028	GO:0003824
Inbred	R	RE	Down	Tabersonine 16-hydroxylase (Fragment) (Cc03_g14110)	-0.00082	0.000015	0.0003	GO:0004497
Hybrid	R	TL	Up	Armadillo/beta-catenin-like repeat ; C2 calcium/lipid-binding domain (CaLB) protein (Cc10_g12190)	0.009225	1.76E-05	<.0001	GO:0005488
Inbred	R	OC	Down	Metal transporter Nramp5 (Cc00_g20930)	-0.00552	0.00021	0.0014	GO:0005215
Inbred	R	OC	Down	Hypothetical protein (Cc00_g07620)	-0.01067	0.000112	0.0001	Several
Inbred	R	RI	Down	Hypothetical protein (Cc00_g07620)	0.000617	2.68E-09	<.0001	Several
Hybrid	R	TH	Up	Phosphate transporter PHO1 homolog 1 (Cc11_g10900)	7.04E-05	9.13E-07	0.0002	Several
Hybrid	R	TH	Up	Phosphate transporter PHO1 homolog 1 (Cc11_g10900)	7.04E-05	9.13E-07	0.0002	Several
Hybrid	R	TH	Up	Putative Uridine-cytidine kinase C (Cc11_g08540)	-0.03415	1.57E-06	<.0001	GO:0000166
Hybrid	R	TH	Up	Putative Uridine-cytidine kinase C (Cc11_g08540)	-0.03415	1.57E-06	<.0001	GO:0000166
Inbred	R	TH	Down	Hypothetical protein (Cc00_g07620)	0.019192	2.35E-06	<.0001	Several
Hybrid	R	RS	Down	Putative uncharacterized protein (Cc00_g03090)	-0.00027	6.48E-07	<.0001	Several
Hybrid	R	RI	Up	Putative pre-mRNA-processing protein 40A (Cc02_g20050)	0.000383	5.62E-07	<.0001	GO:0005515
Inbred	R	RI	Down	Probable trehalose-phosphate phosphatase E (Cc10_g07410)	-3.80E-05	1.44E-09	<.0001	GO:0003824
Inbred	R	TL	Down	Putative uncharacterized protein (Cc00_g03090)	-0.01181	0.001399	0.0011	Several
Hybrid	R	RI	Down	Putative Uncharacterized endoplasmic reticulum membrane protein C16E8.02 (Cc04_g16310)	-7.80E-05	1.18E-09	<.0001	Several
Hybrid	R	TL	Up	Hypothetical protein (Cc00_g07620)	-0.00225	3.23E-05	0.0002	GO:0003678

Appendix 17 continued…

Cultivar	Treat.	Trait	Regulation	Annotation (Gene ID [134])	Estimate	Std Error	Prob>|*t*|	GO term
Hybrid	R	TH	Up	Putative U-box domain-containing protein 50 (Cc07_g14020)	0.001255	6.54E-08	<.0001	GO:0006468
Inbred	R	OC	Down	Benzoate--CoA ligase, peroxisomal (Cc07_g12710)	-0.00379	5.65E-05	0.0002	GO:0003824
Hybrid	R	TL	Down	Pathogenesis-related leaf protein 6 (Cc02_g20280)	-0.00016	1.94E-08	<.0001	GO:0005576
Hybrid	R	RS	Down	Pathogenesis-related leaf protein 6 (Cc02_g20280)	-7.90E-07	3.82E-09	<.0001	GO:0005576
Inbred	R	TH	Down	Thioesterase superfamily protein (Cc04_g11880)	0.008346	2.57E-06	0.0002	Several
Hybrid		RS	Up	Developmentally-regulated GTP-binding protein 1 (Cc07_g16280)	-0.00121	1.59E-06	<.0001	Several
Inbred	R	RE	Down	Putative Probable carboxylesterase 6 (Cc07_g16100)	0.000169	5.47E-06	0.001	GO:0016787
Hybrid	R	RS	Down	EIN3-binding F-box protein 1 (Cc06_g08370)	8.35E-06	1.28E-08	<.0001	Several
Hybrid	R	TH	Down	EIN3-binding F-box protein 1 (Cc06_g08370)	-0.00141	0.000134	0.0018	Several
Inbred	R	RS	Down	1-aminocyclopropane-1-carboxylate synthase 3 (Cc07_g07490)	-0.00085	8.96E-05	<.0001	GO:0003824
Hybrid	R	RS	Up	Hypothetical protein (Cc00_g07620)	-3.70E-05	2.74E-07	<.0001	Several
Inbred	R	TL	Down	Putative methyltransferase DDB_G0268948 (Cc06_g18640)	0.016708	0.002042	0.0012	GO:0008168
Hybrid	R	RS	Up	Putative uncharacterized protein (Cc00_g03090)	-5.40E-05	6.16E-07	<.0001	GO:0003676
Hybrid	R	TH	Up	Golgin candidate 1 (Cc06_g10330)	0.002621	5.38E-07	<.0001	Several
Hybrid	R	RI	Down	PeroxiRedoxin-2B (Cc06_g09990)	-2.60E-05	1.56E-09	<.0001	GO:0045454
Inbred	R	RE	Down	Putative Ankyrin repeat-containing protein At5g02620 (Cc05_g06990)	0.006482	0.000154	0.0006	Several
Hybrid	R	TH	Up	Putative RING finger and CHY zinc finger domain-containing protein 1 (Cc04_g15950)	-0.01405	6.41E-07	<.0001	GO:0008270
Hybrid	R	RI	Down	Dihydrodipicolinate synthase, chloroplastic (Cc08_g04960)	0.00739	8.52E-08	<.0001	GO:0016829